Many Instruments, Weak Instruments, and Microeonometric Practice*

Christian Hansen
Graduate School of Business
University of Chicago

Jerry Hausman
Department of Economics
M.I.T.

Whitney K. Newey
Department of Economics
M.I.T.

September 2004

Abstract

This paper shows that it is important to distinguish between many and weak instrument problems in applications. We find that using Bekker (1994) standard errors that account for many instruments fixes the problems with Angrist and Krueger (1991). We also find that many applications are in a range where this fix is sufficient. To widen the applicability of these standard errors we give theoretical results for non Gaussian models with many and many weak instruments.

JEL Classification:
Keywords: Weak Instruments, Many Instruments, Concentration Parameter.

*The NSF provided partial financial support for this paper.

1 Introduction

The accuracy of asymptotic inference for instrumental variable (IV) estimators is potentially important for microeconometric practice. Weak instruments, where there is low correlation between excluded instruments and endogenous variables, and many instruments, where there is a large number of overidentifying restrictions, can both cause innacuracy of asymptotic approximations. This paper suggests that it is important to distinguish between these problems. In particular, we find that simply using Bekker (1994) standard errors can greatly alleviate the problems in microeconometric applications, a finding like that of Hahn and Inoue (2002).

We assert that, for simplicity's sake, it is important to distinguish between weak instrument and many instrument problems. The many instruments problem has relatively simple solutions. The number of instruments is a choice, and so one solution is just leave some out, as considered in Donald and Newey (2001). Another solution is to use the limited information maximum likelihood (LIML) or Fuller (1977) (FULL) estimators with the Bekker (1994) standard error (BSE). As shown by Morimune (1983) and Bekker (1994), under normality LIML is still approximately normal with many instruments, and the BSE adjust for excess dispersion. In contrast, the weak instrument (WI) problem is more difficult to solve. To obtain WI confidence intervals one has to invert the Kleibergen (2002) or Moreira (2003) test statistics, which is inherently more difficult than just correcting standard errors or dropping insruments.

In this paper we widen the applicability of the BSE by showing that the Bekker (1994) results hold without normality. This extension is important because the normality assumption may often be violated in practice, as we show it to be in the Angrist and Krueger (1991) returns to schooling application. We replace normality by conditional first and second moment assumptions specified below. We also show that FULL is asymptotically equivalent to LIML under Bekker (1994) asymptotics.

To find a way to relax the conditional first and second moment restrictions we also consider asymptotic inference under the many weak instrument asymptotics of Chao and

Swanson (2002) and Stock and Yogo (2003). We obtain the same limiting distribution for LIML and FULL as do Stock and Yogo (2003) under a weaker condition than theirs, that the number of instruments grows slower than the sample size. More importantly, we show that the BSE are consistent under this asymptotics, so that Wald inference can be conducted in the usual way. We also show that the Kleibergen (2003) test statistic is asymptotically correct under many weak instruments.

As pointed out by Hahn, Hausman, and Kuersteiner (2004), FULL provides an alternative to LIML that has much better MSE properties when the instruments are very weak. For this reason we focus on the FULL estimator in this paper, although we give LIML results for comparison purposes. We extend the previous Monte Carlo results by considering confidence interval coverage probabilities.

As evidence that the distinction between weak and many instruments is useful in practice, we offer a fresh analysis of the returns to schooling application of Angrist and Krueger (1991). We find that confidence intervals based on FULL and LIML, with the BSE, give nearly identical results to WI confidence intervals. This finding is consistent with Angrist and Krueger (1991) suffering from a many instrument problem rather than a weak instrument one because, as shown in Bekker and Kleibergen (2003), WI intervals are nearly correct under many instruments, while the BSE do not fix the weak instrument problem. We also find that using all the instruments is more informative than just using a few. We verify these findings in a Monte Carlo study based on the empirical values of parameters, and show that the many instrument approximation is very accurate. These results are consistent with previous Monte Carlo results of Hahn and Inoue (2002).

The reduced form F-statistic (for the instruments excluded from the structural equation) does not distinguish between many instrument and weak instrument problems. If there are many instruments the F-statistic can be arbitrarily small, but FULL and LIML nearly normal, with variability accounted for by the BSE. For this reason we suggest use of the concentration parameter, that is the F-statistic times the number of instruments, for determining whether there is a weak instrument problem. Of course, in the common case of exact identification, the concentration parameter coincides with the F-statistic, but there are a number of applications that are overidentified, including of Angrist and Krueger (1991).

To help determine how large the concentration parameter should be in practice, we give calculation for limits under Staiger and Stock (1997) weak instrument asymptotics, including the BSE. The goal of these calculations is to help determine "cutoff" values for the concentration parameter, where the asymptotics works well above the "cutoffs". We find that for FULL, concentration parameters in the range of 15-20 give coverage probabilities within 10 percent of nominal, with slightly higher values required with more overidentifying restrictions. LIML is close to median unbiased for much smaller values of the concentration parameter, in the range of 5-10, but has less accurate coverage probabilities and wider confidence intervals than FULL. We also find that with many instruments the BSE can give good results even when the F-statistic is very low, e.g. as low as one.

2 Models and Estimators

The model we consider is given by

$$\begin{aligned} \underset{T\times 1}{y} &= \underset{T\times G}{X}\underset{G\times 1}{\delta_0} + \underset{T\times 1}{u}, \\ X &= \underset{T\times K}{Z}\underset{K\times G}{\Pi} + \underset{T\times G}{V}. \end{aligned}$$

For weak and many instrument asymptotic approximations, Z and Π are implicitly allowed to depend on T. Many instruments corresponds to K growing with T and weak instruments to Π shrinking with T. This model differs somewhat from Bekker (1994), in that we assume that the reduced form is correctly specified. For notational convenience we suppress dependence of Z and Π on T.

To describe the estimators let $P_Z = Z(Z'Z)^- Z'$ where A^- denotes any symmetric generalized inverse of a matrix A, i.e. A is symmetric and satifies $AA^-A = A$. We consider estimators of the form

$$\hat{\delta} = (X'P_Z X - \hat{\alpha}X'X)^{-1}(X'P_Z y - \hat{\alpha}X'y).$$

for some choice of α. This class includes all of the familiar k-class estimators, for $\hat{\alpha} = k/(1+k)$, except the least squares estimator. Special cases of these estimators are two-stage least squares (2SLS), where $\hat{\alpha} = 0$, and LIML, where $\hat{\alpha} = \tilde{\alpha}$ and $\tilde{\alpha}$ is the smallest eigenvalue of the matrix $(\bar{X}'\bar{X})^{-1}\bar{X}'P_Z\bar{X}$ for $\bar{X} = [y, X]$. FULL is also a member of this class of estimators, where $\hat{\alpha} = [\tilde{\alpha} - (1-\tilde{\alpha})C/T]/[1-(1-\tilde{\alpha})C/T]$ for some constant C. FULL has moments of all orders, is approximately mean unbiased for $C = 1$, and is second order admissable for $C \geq 4$ under standard large sample, fixed Π asymptotics.

For use in inference we consider the asymptotic variance estimator of Bekker (1994). This variance estimator is consistent under standard asymptotics, many instrument asymptotics, and many weak instrument asymptotics. We give a version of this estimator for any k-class estimator as described above and show its consistency. Let $\hat{u} = y - X\hat{\delta}$ and $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/(T-G)$. The asymptotic variance estimator is given by

$$\hat{\Lambda} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}, \hat{H} = X'P_ZX - \hat{\alpha}X'X,$$
$$\hat{\Sigma} = \hat{\sigma}_u^2[(1-\hat{\alpha})\hat{J} - \hat{\alpha}\hat{H}], \hat{J} = X'P_ZX - \hat{\alpha}X'\hat{u}\hat{u}'X/\hat{u}'\hat{u}.$$

We will show that when normalized appropriately $\hat{\Lambda}$ is a consistent estimator of the asymptotic variance of $\hat{\delta}$. In particular, inference carried out as if $\tilde{\delta}$ were Gaussian with mean δ_0 and variance $\hat{\Lambda}$ is asymptotically correct. It can also be shown that $\hat{\Lambda}$ is identical to the Bekker (1994) estimator of the asymptotic variance of LIMLwhen $\hat{\delta}$ is the LIML estimator.

3 Quarter of Birth and Returns to Schooling

A much analyzed microeconometric application is Angrist and Krueger's (1991) study of the returns to schooling using quarter of birth as an instrument. Bound, Jaeger, and Baker (1996) asserted that this application suffered from weak instrument problems, which helped to motivate much of the theoretical econometric work. For example, Staiger and Stock (1997) considered this study as their main application. Thus, it seems appropriate to begin our discussion with this application. The data we consider is the 1930-1939 cohort, as considered in Donald and Newey (2001).

Figures 1-5 are graphs of confidence intervals at different significance levels using several different methods. The confidence intervals we consider are based on Two-Stage Least Squares (2SLS) with the usual (asymptotic) standard errors, LIML with usual standard errors, LIML with the Bekker (1994) standard errors that are robust to many instruments, and the Kleibergen (2002) confidence interval which is robust to weak instruments. Results for the Fuller estimator give similar answers to LIML.

Figure 1 shows that with 3 excluded instruments (two overidentifying restrictions), 2SLS and WI confidence intervals using Kleibergen (2002) are very similar. The main difference seems to be a slight horizontal shift. Since the WI confidence intervals are centered about the LIML estimator, this shift corresponds to a slight difference in the LIML and 2SLS estimators. This difference is consistent with 2SLS having slightly higher bias than LIML. This result is consistent with no weak instrument problem for 3 instruments, which was asserted by Bound, Jaeger, and Baker (1996). Figure 2 shows that with 180 excluded instruments (179 overidentifying restrictions) the confidence intervals are quite different. In particular, there is a much more pronounced shift in the 2SLS location, as well as smaller dispersion. These results are consistent with a larger bias in 2SLS resulting from the many instruments.

Figure 3 compares the confidence interval for LIML based on the usual standard error formula for 180 instruments with the WI confidence interval. Here we find that the WI interval is wider than the usual one. Figure 4 compares the WI interval with one based on the BSE. Here we find that the WI interval is nearly identical to the BSE one. Thus, using Bekker (1994) standard errors has the same effect as using the WI interval in this application.

It is also interesting to compare Figures 1 and 4. Because there are on the same scale, it is easy to see that the confidence intervals in Figure 4 are narrower than those in Figure 1. In this sense using all 180 instruments turns out to be more informative in this application than just using 3. This is also consistent with the results shown in Figure 5, which gives BSE and WI confidence intervals for the 180 instrument case when age and age squared are added as covariates. Again, the confidence intervals are almost identical. Also, even with age and age squared added as covariates, we find that the data are informative about returns to schooling when 180 instruments are used, and that many instrument confidence intervals are similar to the weak instrument ones. Cruz and Moreria (2003) found similar WI intervals in this application using Moreira's (2003) conditional likelihood ratio method.

It is interesting to note that following Donald and Newey's (2001) approach would also lead to the conclusion that using all of the instruments was informative, if Bekker standard errors were used for the LIML estimator. They found that with 2SLS a mean square error criteria chooses just 3 instruments. They also found that with LIML a mean-suare error criteria chooses all 180 instruments. The 2SLS estimator with 3 instruments is .1077 with standard error .0195. The LIML estimator with 180 instruments is .1089 with BSE .0166. Thus, the estimated returns to schooling are similar for 2SLS with 3 instruments and for LIML with 180 instruments, but the BSE are smaller for the 180 instruments, suggesting using all the instruments is more informative.

In summary, we find that with a few instruments the WI and usual standard errors give very similar results, while with many instruments the WI and BSE intervals are very similar. As shown in Bekker and Kleibergen (2003), WI intervals based on Kleibergen (2002) provide a bound on many instrument confidence intervals which is very tight when there are few instruments relative to the sample size. Furthermore, there is no reason to think that if there were weak instruments the BSE and WI intervals would be similar. Thus, Figures 1-5 are consistent with there being a many instrument problem in this application rather than a weak instrument problem.

More convincing evidence is provided by a Monte Carlo study of the properties of the estimators based on the full sample. We carried out two experiments, one where the reduced form had three excluded instrumental variables, and one where the reduced form had 180 excluded instrumental variables. The disturbances were chosen to be Gaussian, and the parameters of the model to correspond to the estimates obtained from the application. Table 1 reports the results of this experiment, giving bias, mean-square error, and coverage probability for nominal 95 percent confidence intervals based on the estimator and the usual standard error. We find that with 3 excluded instruments all of the estimators perform well, including 2SLS. We also find that with 180 instruments, the coverage of the standard 2SLS and LIML confidence intervals is quite poor, but that with Bekker standard errors the LIML confidence intervals are just right. Thus, in this Monte Carlo study we find evidence that using the Bekker (1994) standard errors takes care of whatever inference problem might be present in this data.

Table 1
Males born 1930-1939. 1980 IPUMS
Full sample. n =329,509, $\beta = .0953$

	Bias/Beta	MSE*100	Size
3 instruments, $\Pi' zz' \Pi/\sigma^2 = 98.60$			
2SLS	-0.0061	0.0420	0.051
LIML	0.0011	0.0440	
se1			0.056
Bekker			0.052
Fuller 1	-0.0027	0.0430	
Bekker			0.049
Kleibergen			0.056
180 instruments, $\Pi' zz' \Pi/\sigma^2 = 436.98$			
2SLS	-0.0988	0.0151	0.221
LIML	0.0012	0.0127	
se1			0.155
Bekker			0.050
Fuller 1	0.0005	0.0126	
Bekker			0.051
Kleibergen			0.051

In summary, we find that using Bekker (1994) standard errors fixes the inference problem in the Angrist and Krueger (1991) application, and produces confidence intervals that are very similar to weak instrument intervals. These results are consistent with the usual inference methods being incorrect because of a many intstruments problem rather than a weak instruments problem.

4 The Concentration Parameter Versus the F-Statistic

To reconcile our results with those of the previous literature, we make the claim that the concentration parameter provides a better measure of whether instruments are weak than does the F-statistic. In the case of one endogenous variable and no covariates, an estimate of the concentration parameter is $nR^2/(1 - R^2)$, where R^2 is the r-squared of the reduced form. The F-statistic is this divided by the number of instruments. Under the many instrument asymptotics of Morimune (1983) and Bekker (1994), the F-statistic will converge to a constant, that can be arbitrarily small. Thus, even though LIML is becoming normally distributed, the F-statistic could indicate weak instruments. In contrast, the concentration parameter goes to infinity under many instrument asymptotics, and thus seems a better measure of how well the asymptotic approximation is working.

The central role of the concentration parameter in determining the accuracy of the asymptotic approximation has been well understood since the early work in simultaneous equations; see Rothenberg (1984). Our purpose is to determine just how large the concentration parameter needs to be to obtain accurate coverage probabilities and tight confidence intervals for the FULL and LIML estimators with the BSE. We will show that the BSE account well for the presence of overidentifying restrictions.

For these purposes we consider the weak instrument limit of the FULL and LIML estimators and t-ratios under the Staiger and Stock (1997) asymptotics. This limit is obtained by letting the sample size to go infinity while holding the concentration parameter fixed. As shown in Staiger and Stock (1997), it provides excellent approximations to small sample distributions. Furthermore, it seems very appropriate for microeconometric settings, where the sample size is often quite large relative to the concentration parameter.

Tables 2 and 3 give results for the median, interquartile range, and coverage probability of nominal 5 percent confidence intervals based on BSE for LIML and FULL, with $C = 1$ for FULL, for a range of number of instruments K and concentration parameter mu^2. The value of rho is set to .5 throughout these tables. Table 2 gives results for several different numbers of instruments. Table 3 focus on the exactly identified case. We find that LIML is nearly (median) unbiased for small values of the concentration parameter, but that it has large dispersion. In Table 3, once the concentration parameter reaches 5 the coverage probabilities for the LIML estimator are .4 or above, but they converge slowly to their nominal value as the concentration parameter rises. FULL is much less dispersed for small concentration parameters than is LIML but more biased. Its coverage probability approaches the nominal value faster, but is further away for small values of the concentration parameter. Overall with K=1 we see quite good results for LIML once the concentration parameter reaches about 10, with FULL being less dispersed always and having a more accurate coverage probability once the parameter reaches 15. Table 2 shows similar results except that the concentration parameter needs to be larger for good results the more instruments there are. For example, with K=4 LIML is nearly unbiased with a good coverage probability with a concentration parameter of about 8. Also, in Table 2, letting the number of instruments grow at the same rate as the concentration parameter leads to accurate asymptotic approximations even when the raio (i.e. the F-statistic) is small. For example, with a ratio mu^2/K=1, the asymptotic approximaiton is quite accurate at K=16, a remarkably small value.

As further evidence we offer the results of a few Monte Carlo studies. Tables 4, 5, and 6 show results when the concentration parameter is 10, 20, and 35 respectively, with a sample size of 100, for several different numbers of instruments, and for correlation coefficient equal to .4 and .6. Results are not as good as in Tables 2 and 3, which may be do to the very small sample size. With a concentration parameter of 10 (in Table 4), the asymptotic approximation is not very accurate, even for the one instrument case. In particular, we find that the bias of LIML increases substantially as the number of instruments increase. However, for concentration parameter 35 the asymptotic approximation is quite good, even for large numbers of instruments. In particular, the bias of LIML increases very little with the number of instruments. Also, the use of Bekker (1994) standard errors makes the coverage probabilities for LIML confidence intervals remarkably stable across the number of instruments.

We also carried out Monte Carlo work for larger values of the concentration parameter. For brevity, we do not report results here. They all give improvements over the results for a concentration parameter of 35, which are quite good. These Monte Carlo results also show how the F-statistic can give a misleading picture concerning the accuracy of the asymptotic approximation. With a concentration parameter of 20 and 10 instruments, the expected value of the F-statistic would be about 2, which is in the range that would be widely considered in the literature to give a weak instrument problem. Nevertheless, the LIML estimator with Becker standard errors performs quite well in the case, having a bias of less than .01 and having nominal 5% coverage probabilities of .432 for $\rho = .4$ and .571 for $\rho = .6$.

These results are also consistent with recent Monte Carlo work of Davidson and MacKinnon (2004). From careful examination of their graphs it appears that with few instruments the bias of LIML is very small once the concentration parameter exceeds 10, and that the variance of LIML is quite small once the concentration parameter exceeds 20.

Here we have focused on LIML and FULL with the BSE because, in the overidentified case, 2SLS tends to be much more poorly behaved. In particular, its bias increases very rapidly with the number of overidentifying restrictions, as is suggested by the higher-order asymptotics. Partly as a result, the coverage probability for 2SLS confidence intervals also deterioate rapidly. Also, as we will discuss below, there is theoretical reason to expect LIML to have lower bias than 2SLS in the case of weak instruments. For these reasons we prefer FULL LIML to 2SLS in the weak instrument setting.

Returning now to the quarter of birth, returns to schooling application, it is interesting to note that with 3 excluded instruments the concentration parameter, that is three times the F-statistic, is equal to 98.60. With 180 instruments the concentration parameter is 436.98. Both of these are well up in the range where we would expect good performance of FULL and LIML with Bekker standard errors, as we found in the Monte Carlo results reported above. In contrast, with 180 instruments, the F-statistic would be less than 3, which is in the range that was considered weak instruments by Bound, Jaegar and Baker (1996).

To consider wider implications for empirical practice we have examined some of the existing empirical studies that use instrumental variables and report reduced form results. Table 7 gives the value of the concentration parameter from empirical studies for the last five years of the AER, JPE, and QJE.

Empirical Results: All papers last five years, three journals.

	Num Papers	**Median**	**Q10**	**Q25**	**Q75**	**Q90**
K	57	2	1	2	4.25	8
mu^2	28	23.6	8.95	12.7	105	588
rho	22	.279	.022	.0735	.466	.555

several studies. Here most of the concentration parameters seem to be in a range where the BSE would work well. From these results we draw the tentative conclusion that for many of the studies we consider the instruments are strong enough so that the usual asymptotic approximation for FULL and LIML, although there are a number of exceptions in the lower tail of the concentration parameter distribution.

5 Improved Inference with Weak Instruments

As discussed in Staiger and Stock (1997), the weak instrument limit of the 2SLS and LIML estimators are the finite sample distributions of the 2SLS and LIMLK estimator, where the sample size is equal to the square root of the concentration parameter. Consequently, asymptotic results that have been shown for 2SLS and LIML hold under weak instruments as the concentration parameter grows. In particular, as the concentration parameter grows, the LIML limit will have less median bias than the 2SLS. Also, as the concentration parameter grows at the same rate as the number of instruments, the ratio of the LIML estimator to the BSE will converge to a standard normal.

Under weak instruments, an unbiased estimator of the concentation parameter is given by

$$Y'Z(Z'Z)^{-1}Z'Y/\hat{\sigma}_V^2 - K.$$

This provides an alternative way to estimate the concentration parameter that should be more accurate and may be useful knowing when to use a weak instrument approximation. Some Monte Carlo work on such a procedure was done. It leads to accurate results, except when the true concentration parameter is around the cutoff value for using weak instruments.

One approach to improved approximation is to use the weak instrument limit, plugging in estimates of the concentration parameter and rho, to obtain critical values for confidence intervals and tests. One can show, using the usual bootstrap arguments, that this gives an improved approximation as the concentration parameter grows (with either fixed number of instruments or the number growing at the same rate). The rapidity with which the asymptotic approximation works well as the concentration parameter grows suggests this approach might be useful. We leave examination of this approach to the next version of the paper.

6 Many Instrument Asymptotics Without Normality

One highlight of the results so far is the accuracy imparted to the confidence intervals by the BSE. One problem for applying the results of Bekker (1994) in practice is that they assume the disturbances are Gaussian, i.e. normally distributed. This assumption may often be violated in practice. For example, Figure 6 gives a plot of a kernel estimator of the structural residual density from 2SLS and a plot of a Gaussian density with the same mean and variance as the kernel density. We find that the actual residual density is thicker tailed and more peaked than the Gaussian.

To help widen the field of application of many instrument asymptotics, we give here results that allow for non-Gaussian disturbances. We first consider many instrument asymptotics, where the number of instruments K grows at the same rate as the sample size. We will also consider many weak instrument asymptotics, where K grows slower than T, but K grows at the same rate as the degree of identification.

For the many instrument asymptotics we will consider Z as nonrandom. Alternatively, one could interpret the following results as being conditional on Z. We will make use of two assumptions, the first of which is for consistency and the second of which is added for asymptotic normality. Let Z_t, u_t, V_t denote the t^{th} row of Z, u, and V respectively.

Assumption 1: $(u_1, V_1), ..., (u_T, V_T)$ are i.i.d. with mean zero and finite fourth moments, the variance of (u_t, V_t) is nonsingular, $Z'Z$ is nonsingular, and as $T \longrightarrow \infty$ there is a scalar α with $0 < \alpha < 1$ and a positive definite matrix Q such that

$$K/T \longrightarrow \alpha, \Pi' Z' Z \Pi / T \longrightarrow Q.$$

This condition allows the number of instruments to grow at the same rate as the sample size, but requires that $\Pi' Z' Z \Pi / T$ converges. In this sense adding additional instruments does not add information. The restriction that $Z'Z$ is nonsingular is essentially a normalization. Alternatively, we could interpret K as the rank of $Z'Z$. For the second assumption, let $\sigma_{Vu} = E[V_t u_t]$, $\sigma_u^2 = E[u_t^2]$, $\gamma = \sigma_{Vu}/\sigma_u^2$, and $\tilde{V} = V - u\gamma'$, having t^{th} row $\tilde{V}_t'$.

Assumption 2: $E[u_t|\tilde{V}_t] = 0$, $E[u_t^2|\tilde{V}_t] = \sigma_u^2$, for some $p > 2$, $E[|u_t|^p|\tilde{V}_t]$ is bounded, and $\max_{t \leq T} \|\Pi Z_t'\|/\sqrt{T} \longrightarrow 0$.

The vector $\tilde{V}_t$ consists of residuals from the population regression of V_t on u_t and so satisfies $E[\tilde{V}_t u_t] = 0$ by construction. Under joint normality of (u_t, V_t), u_t and $\tilde{V}_t$ are independent, so the first two conditions automatically hold. In general these two conditions weaken the joint normality restriction to first and second moment independence of u_t from $\tilde{V}_t$. The other two conditions are useful for the central limit theorem, with the last one implying asymptotic normality of $\Pi' Z' u/\sqrt{T}$. It is interesting to note that no other restrictions are imposed on Z.

The following is a consistency result for the class of estimators described in Section 2.

THEOREM 1: *If Assumption 1 is satisfied and* $\hat{\alpha} \xrightarrow{p} \alpha$ *then* $\hat{\delta} \xrightarrow{p} \delta_0$. *Also, for LIML,* $\hat{\alpha} \xrightarrow{p} \alpha$.

To interpret the condition $\hat{\alpha} \xrightarrow{p} \alpha$, note that the estimator $\hat{\delta}$ satisfies

$$\hat{\delta} = \delta_0 + (\frac{X' P_Z X}{T} - \hat{\alpha}\frac{X'X}{T})^{-1}(\frac{X' P_Z u}{T} - \hat{\alpha}\frac{X'u}{T}). \qquad (6.1)$$

With fixed instruments $X'P_Z u/T \xrightarrow{p} 0$. With many instrument, this no longer holds. By a standard calculation we know that for scalar X_t,

$$E[X'P_Z u] = E[V'P_Z u] = trE[P_Z u'V] = \sigma_{Vu} tr(P_Z) = \sigma_{Vu} rank(Z). \tag{6.2}$$

Then, because $rank(Z) = K$ grows at the same rate as T, $E[X'P_Z u/T]$ does not shrink to zero as T grows. In fact, it turns out that $X'P_Z u/T \xrightarrow{p} \alpha\sigma_{Vu}$,Therefore, for $\hat{\alpha} \xrightarrow{p} \alpha$,

$$\frac{X'P_Z u}{T} - \hat{\alpha}\frac{X'u}{T} \xrightarrow{p} \alpha\sigma_{Vu} - \alpha\sigma_{Vu} = 0,$$

leading to consistency. The LIML result is explained by the fact that $\tilde{\alpha} = \hat{u}'P_Z\hat{u}/\hat{u}'\hat{u}$. When the numerator and denominator are each divided by T, they will converge to $\alpha\sigma_u^2$ and σ_u^2 repectively, so that $\tilde{\alpha} \xrightarrow{p} \alpha$.

The next result shows asymptotic normality of the estimators under many instrument asymptotics. Let $\tilde{\Omega} = E[\tilde{V}_t\tilde{V}_t']$.

THEOREM 2: *If Assumptions 1 and 2 are satisfied and $\hat{\alpha} = \tilde{\alpha} + o_p(1/\sqrt{T})$ then*

$$\sqrt{T}(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda), T\hat{\Lambda} \xrightarrow{p} \Lambda,$$
$$\Lambda = \sigma_u^2 Q^{-1} + \sigma_u^2 \frac{\alpha}{1-\alpha} Q^{-1}\tilde{\Omega}Q^{-1}.$$

We find that, as in Bekker (1994), under many instruments the asymptotic variance of LIML is larger than the asymptotic variance $\sigma_u^2 Q^{-1}$ under fixed instrument asymptotics. We also find that any of the estimators with $\hat{\alpha}$ asymptotically close to the LIML $\tilde{\alpha}$ will be asymptotically normal with the same asymptotic variance as LIML. In particular, for FULL,

$$\hat{\alpha} - \tilde{\alpha} = -[(1-\tilde{\alpha})^2 C/T]/[1 - (1-\tilde{\alpha})C/T] = o_p(1/\sqrt{T}),$$

so that FULL has the same asymptotic variance as LIML under many instrument asymptotics.

We can also compare the asymptotic variance of the LIML estimator with another estimator $\breve{\delta}$ based on an explicity bias correction with $\hat{\alpha} = K/T$. Note that

$$\sqrt{T}(\breve{\delta} - \delta_0) = (X'(P_Z - \frac{K}{T})X/T)^{-1}X'(P_Z - \frac{K}{T})u/\sqrt{T}.$$

Similarly to the derivations in the Appendix,

$$X'(P_Z - \frac{K}{T})X/T \xrightarrow{p} Q + \alpha\Omega - \alpha(Q + \Omega) = (1 - \alpha)Q.$$

Assume that $\sqrt{T}(K/T - \alpha) \longrightarrow 0$, so that $(K/T - \alpha)X'u/\sqrt{T} \xrightarrow{p} 0$, we have

$$\begin{aligned} X'(P_Z - \frac{K}{T})u/\sqrt{T} &= X'(P_Z - \alpha I)u/\sqrt{T} + o_p(1) \\ &= (1-\alpha)\Pi'Z'u/\sqrt{T} + V'(P_Z - \alpha I)u/\sqrt{T} + o_p(1) \\ &= (1-\alpha)\Pi'Z'u/\sqrt{T} + \tilde{V}'(P_Z - \alpha I)u/\sqrt{T} \\ &\quad + \gamma u'(P_Z - \alpha I)u/\sqrt{T} + o_p(1). \end{aligned}$$

The sum of first two terms following the last equality are identical to those that appear in the LIML derivation, so that the variance of their sum will converge to Σ. Furthermore, the last term will be uncorrelated with each of those terms if third moments of u_t are zero conditional on $\tilde{V}_t$. Therefore, under these conditions the asymptotic variance of $X'(P_Z - K/T)u/\sqrt{T}$ will be larger than the corresponding one for LIML, and hence the asymptotic variance of the bias corrected estimator larger than that of LIML. For many weak instrument asymptotics, Chao and Swanson (2004) carry out an analogous efficiency comparison, finding that LIML is efficient relative to a wide class other estimators when the disturbances have an elliptically symmetric distribution.

An alternative approach to accounting for many instruments is asymptotics like that of Chao and Swanson (2002), where the number of instruments grows slower than the sample size but the concentration parameter grows at the same rate as the number of instruments. This type of aproximation also seems well suited to microeconometric applications, where often identification is not very strong and the sample size is very large. We show asymptotic normality of k-class estimators when $K/T \longrightarrow 0$. We also show that the BSE remains consistent under this asymptotics and that the Kleibergen (2003) confidence intervals are asymptotically correct under this asymptotics.

For this asymptotics it is convenient to switch to the case where Z_t is i.i.d. along with u_t and V_t. The following assumption imposes this and other conditions for consistency.

Assumption 3: $(u_1, V_1, Z_1), ..., (u_T, V_T, Z_T)$ are i.i.d. with finite fourth moments, $E[(u_t, V_t)|Z_t] = 0$, $Var((u_t, V_t)|Z_t)$ is constant and nonsingular, for a positive constant C, $E[u_t^4|Z_t] \leq C$, $E[\|V_t\|^4|Z_t] \leq C$, $E[Z_tZ_t']$ is nonsingular, $K/T \longrightarrow 0, TE\left[\|\Pi' Z_t\|^4\right]/K^2 \longrightarrow 0$, $E[\{Z_t'(E[Z_tZ_t'])^{-1}Z_t\}^2]/KT \longrightarrow 0$, and there is a nonsingular matrix $\tilde{Q}$ with

$$T\Pi' E[Z_tZ_t']\Pi/K \longrightarrow \tilde{Q}. \tag{6.3}$$

Equation (6.3) is the critical condition leading to many weak instruments asymptotics. An example is given by scalar X_t, $E[Z_tZ_t'] = I_K$, and $\Pi = B(1/\sqrt{T}, ..., 1/\sqrt{T})'$. In this case $T\Pi' E[Z_tZ_t']\Pi/K = B^2$. Here each reduce form coefficient goes to zero at the same rate as $1/\sqrt{T}$, that is like weak instruments, but the number of instruments grows with the sample size.

Assumption 3 also includes rate conditions. If the elements of Z_t are uniformly bounded under the normalization $E[Z_tZ_t'] = I_K$, so that $\|Z_t\|^2 \leq CK$ for a constant C, then $K/T \longrightarrow 0$ will suffice for these conditions. To see this, note that $\|\Pi\|^2 = O(K/T)$, so that

$$\begin{aligned} TE\left[\|\Pi' Z_t\|^4\right]/K^2 &\leq CT\|\Pi\|^2 E[\|\Pi' Z_t\|^2]/K = CT\|\Pi\|^4/K = O(K/T) \longrightarrow 0, \\ E[(Z_t'Z_t)^2]/KT &\leq CK^2/KT = CK/T \longrightarrow 0. \end{aligned}$$

For asymptotic normality we add another rate condition.

Assumption 4: $tr(E[\|Z_t\|^2 Z_tZ_t']^2)/K^2T \longrightarrow 0$.

For bounded Z_t, this rate conditions will also be satisfied when $K/T \longrightarrow 0$, since then

$$tr(E[\|Z_t\|^2 Z_tZ_t']^2)/K^2T \leq Ctr(I_K^2)/T \longrightarrow 0.$$

Assumption 3 will suffice for consistency.

THEOREM 3: *If Assumption 3 is satisfied and* $T\hat{\alpha}/K \xrightarrow{p} 1$ *then* $\hat{\delta} \xrightarrow{p} \delta_0$. *Also, for LIML,* $T\hat{\alpha}/K \xrightarrow{p} 1$.

The condition that $T\hat{\alpha}/K \xrightarrow{p} 1$ is analogous to the condition $\hat{\alpha} \xrightarrow{p} \alpha$ from Theorem 1. The estimator $\hat{\delta}$ satisfies

$$\hat{\delta} = \delta_0 + (\frac{X'P_Z X}{K} - (T\hat{\alpha}/K)\frac{X'X}{T})^{-1}(\frac{X'P_Z u}{K} - (T\hat{\alpha}/K)\frac{X'u}{T}). \tag{6.4}$$

We know by the law of large numbers that $X'u/T \xrightarrow{p} \sigma_{Vu}$, and it also turns out that, similarly to equation (6.2), $X'P_Z u/K \xrightarrow{p} \sigma_{Vu}$, so the right hand side should converge to zero under $T\hat{\alpha}/K \xrightarrow{p} 1$. Also, under Assumption 3, $X'P_Z X/K - X'X/T \xrightarrow{p} \tilde{Q}$, leading to consistency of $\hat{\delta}$.

The next result is asymptotic normality under many weak instruments.

THEOREM 4: *If Assumptions 3 and 4 are satisfied and* $\hat{\alpha} = \tilde{\alpha} + o_p(\sqrt{K}/T)$ *then*

$$\sqrt{K}(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \tilde{\Lambda}), K\hat{\Lambda} \xrightarrow{p} \tilde{\Lambda},$$
$$\tilde{\Lambda} = \sigma_u^2 \tilde{Q}^{-1} + \sigma_u^2 \tilde{Q}^{-1}\tilde{\Omega}\tilde{Q}^{-1}.$$

This result does not impose any restrictions on the conditional distribution of u_t given $\tilde{V}_t$, unlike Theorem 2, and is more general in that sense.

The approximations to the variance of $\tilde{\delta}$ from Theorems 2 and 4 are quite close to each other when K/T is small. If we consider $\Pi' E[Z_t Z_t']\Pi \approx \Pi' Z' Z \Pi/T$, meaning the expressions are close, and $\alpha \approx K/T$ then $\tilde{Q} \approx (T/K)Q$, so that

$$\tilde{\Lambda}/K \approx \sigma_u^2 Q^{-1}/T + (K/T)\sigma_u^2 Q^{-1}\tilde{\Omega}Q^{-1}/T,$$
$$\Lambda/T \approx \sigma_u^2 Q^{-1}/T + [(K/T)/(1 - K/T)]\sigma_u^2 Q^{-1}\tilde{\Omega}Q^{-1}/T.$$

Thus, the variance approximations for $\tilde{\delta}$ for the Bekker (1994) and many weak instrument asymptotics differ only by the factor $1/(1 - K/T)$ in the second term. When the number of instruments is small relative to the sample size this factor will be very close to 1, so the variance approximations will differ little. For example, for the Angrist and Krueger application, where K is 180 and T is over 320,000, the differences in the approximate variances are tiny.

The Kleibergen (2003) confidence intervals turn out to be asymptotically correct under the weak instrument asymptotics. In contrast, under Bekker (1994) asymptotics the

Kleibergen confidence intervals provide only bounds, as shown by Bekker and Kleibergen (2003). It does turn out that the bounds are nearly sharp when K/T is close to zero. The intuition is that the Kleibergen confidence intervals are only wrong under Bekker (1994) asymptotics because of the extra term, which is close to one in most applications.

To state the result, let $u(\delta) = y - X\delta$, $\tilde{X}(\delta) = X - u(\delta)[u(\delta)'X/u(\delta)'u(\delta)]$, and

$$\hat{S}(\beta) = Tu(\delta)'P_Z\tilde{X}(\delta)[\tilde{X}(\delta)'P_Z\tilde{X}(\delta)]^{-1}\tilde{X}(\delta)'P_Z u(\delta)/u(\delta)'u(\delta).$$

THEOREM 5: *If Assumptions 3 and 4 are satisfied then* $\hat{S}(\beta_0) \xrightarrow{d} \chi^2(p)$.

Because of this one can form joint confidence intervals for the vector β_0 by inverting the Kleibergen (2002). However, since we have asymptotic normality and a consistent estimator of the asymptotic variance, it is simpler to just proceed with Wald type inference in the usual way.

7 Appendix: Proofs of Theorems.

Throughout, let C denote a generic positive constant that may be different in different uses.

LEMMA A1: If (u_i, v_i, z_i) are independent with $E[u_i|z_i] = E[v_i|z_i] = 0$, $E[u_i^4|z_i] \leq C$, $E[v_i^4|z_i] \leq C$, z_i is $K \times 1$, then for $P_z = Z(Z'Z)^{-}Z'$ and $\tilde{Z} = (z_1, ..., z_n)$,

$$u'P_Z v - E[u'P_Z v|\tilde{Z}] = O_p(\sqrt{K}).$$

Proof: Let $\sigma_{uvi} = E[u_i v_i|z_i]$, $\mu_{ui}^j = E[(u_i)^j|z_i]$, $\mu_{vi}^j = E[(v_i)^j|z_i]$. By i.i.d. data, $E[uv'|\tilde{Z}] = diag(\sigma_{uv1}, ..., \sigma_{uvn}) = \Gamma$. Then

$$E[u'P_Z v|\tilde{Z}] = tr(P_Z E[vu'|\tilde{Z}]) = tr(P_Z\Gamma).$$

Also, for $P_Z = [p_{ij}]_{i,j=1}^n$,

$$E[(u'P_Zv)^2|\tilde{Z}] \tag{7.5}$$
$$= \sum_{i,j,k,\ell=1}^n p_{ij}p_{k\ell}E[u_iv_ju_kv_\ell|\tilde{Z}] = \sum_{i=1}^n p_{ii}^2 E[u_i^2v_i^2|z_i] + \{\sum_{i\neq j=1}^n (p_{ii}p_{jj}+p_{ij}^2)\sigma_{uvi}\sigma_{uvj} + p_{ij}^2\mu_{ui}^2\mu_{vi}^2\}$$
$$= \sum_{i=1}^n p_{ii}^2\{E[u_i^2v_i^2|z_i] - 2\sigma_{uvi}\sigma_{uvj} - \mu_{ui}^2\mu_{vj}^2\} + tr(P_Z\Gamma)^2 + \sum_{i,j=1}^n p_{ij}^2(\sigma_{uvi}\sigma_{uvj} + \mu_{ui}^2\mu_{vj}^2)$$
$$\leq C\sum_{i=1}^n p_{ii}^2 + C\sum_{i,j=1}^n p_{ij}^2 + tr(P_Z\Gamma)^2 \leq C\sum_{i,j=1}^n p_{ij}^2 + tr(P_Z\Gamma)^2$$

We have

$$\sum_{i,j=1}^n p_{ij}^2 = \sum_{i=1}^n Z_i'(Z'Z)^-(\sum_{j=1}^n Z_j'Z_j)(Z'Z)^-Z_i$$
$$= tr((Z'Z)^-Z'Z(Z'Z)^-Z'Z) = tr((Z'Z)^-Z'Z) \leq K.$$

Then equation (7.5) gives

$$E[(u'P_Zv - E[u'P_Zv|\tilde{Z}])^2|\tilde{Z}] \leq C\sum_{i,j=1}^n p_{ij}^2 \leq CK,$$

so the conclusion follows by the conditional Markov inequality. Q.E.D.

LEMMA A2: *If Assumption 1 is satisfied then for* $\bar{\Omega} = Var(u_t + V_t'\delta_0, V_t)$, $\bar{Q} = F = \begin{bmatrix} \delta_0'Q\delta_0 & \delta_0'Q \\ Q\delta_0 & Q \end{bmatrix}$,

$$\bar{X}'\bar{X}/T \xrightarrow{p} \bar{\Omega} + \bar{Q}, \bar{X}'P_Z\bar{X}/T \xrightarrow{p} \alpha\bar{\Omega} + \bar{Q}.$$

Proof: First, for $\bar{V} = [u + V\delta_0, V], \bar{V}'\bar{V}/T \xrightarrow{p} \bar{\Omega}$ holds by Khintchines law of large numbers. Also, for $\bar{\Pi} = [\Pi\delta_0, \Pi]$ we have $\bar{\Pi}'Z'Z\bar{\Pi}/T \longrightarrow \bar{Q}$. Also, for $\bar{V}_j$ equal to the j^{th} column of $\bar{V}$, $E[\bar{\Pi}'Z'\bar{V}_j/T\left(\bar{\Pi}'Z'\bar{V}_j/T\right)'] = \bar{\Omega}_{jj}\bar{\Pi}'Z'Z\bar{\Pi}/T^2 \longrightarrow 0$, we have $\bar{\Pi}'Z'\bar{V}/T \xrightarrow{p} 0$. The first conclusion then follows from

$$\bar{X}'\bar{X}/T = (Z\bar{\Pi} + \bar{V})'(Z\bar{\Pi} + \bar{V})/T.$$

and the triangle inequality. Also, note that $E[\bar{V}'P_Z\bar{V}/T] = (K/T)\bar{\Omega}$, so by Lemma A1, applied to each element of the matrix $\bar{V}'P_Z\bar{V}/T$, we find that $\bar{V}'P_Z\bar{V}/T - \alpha\bar{\Omega} =$

$O_p(\sqrt{K}/T) + (K/T - \alpha)\bar{\Omega} \xrightarrow{p} 0$. The second conclusion then follows from the triangle inequality and

$$\bar{X}'P_Z\bar{X}/T = \bar{V}'P_Z\bar{V}/T + Z\bar{\Pi}'\bar{V}/T + \bar{V}'Z\bar{\Pi}/T + \bar{\Pi}'Z'Z\bar{\Pi}/T. Q.E.D.$$

Proof of Theorem 1: By Lemma A2, when $\hat{\alpha} \xrightarrow{p} \alpha$ we have

$$\frac{X'P_ZX}{T} - \hat{\alpha}\frac{X'X}{T} \xrightarrow{p} \alpha\Omega + Q - \alpha(\Omega + Q) = (1-\alpha)Q,$$

which is nonsingular by Assumption 1. Also, by Lemma A1,

$$V'P_Zu/T = E[V'P_Zu/T] + O_p(\sqrt{K}/T) = (K/T)\sigma_{Vu} + o_p(1) \xrightarrow{p} \alpha\sigma_{Vu}$$

and $X'u/T \xrightarrow{p} \sigma_{Vu}$ also holds. The first conclusion then follows from eq. (6.1) and the continuous mapping theorem.

To prove the second conclusion, note that for $a > 0$ and $b \geq 0$, $(\alpha a+b)/(a+b) \geq \alpha$,with equality if $b = 0$. By Assumption 1 $\bar{\Omega}$ is nonsingular, so for any vector ζ with $\|\zeta\| = 1$ we have

$$R(\zeta) = \frac{\zeta'(\alpha\bar{\Omega} + \bar{Q})\zeta}{\zeta'(\bar{\Omega} + \bar{Q})\zeta} \geq \alpha,$$

with equality if and only if $\zeta'\bar{Q}\zeta = 0$. Let $\hat{R}(\zeta) = \zeta'\bar{X}'P_Z\bar{X}\zeta/\zeta'\bar{X}'\bar{X}\zeta$. By Lemma A2 $\hat{R}(\zeta) \xrightarrow{p} R(\zeta)$ and this convergence can be shown to be uniform in ζ with $\|\zeta\| = 1$ by standard arguments. Also, for $\zeta^* = (1, -\delta_0')'/\left\|(1, -\delta_0')\right\|$, we have $\bar{Q}\zeta^* = 0$, so that $R(\zeta^*) = \alpha$. Furthermore, by standard matrix results $\tilde{\mu} = \min_{\|\zeta\|=1} \hat{R}(\hat{\zeta})$ is the smallest root μ of $\bar{X}'P_Z\bar{X}\zeta = \mu\bar{X}'\bar{X}\zeta$. Multiplying through by $(\bar{X}'\bar{X})^{-1}$, $\tilde{\mu}$ will also be the smallest root of the equation $(\bar{X}'\bar{X})^{-1}\bar{X}'P_Z\bar{X}\zeta = \mu\zeta$,i.e. the smallest eigenvalue of the matrix $(\tilde{X}'\tilde{X})^{-1}\tilde{X}'P_Z\tilde{X}$, so $\tilde{\alpha} = \min_{\|\zeta\|=1} \hat{R}(\hat{\zeta})$. It then follows by standard arguments that

$$\tilde{\alpha} - \alpha = \min_{\|\zeta\|=1} \hat{R}(\zeta) - \min_{\|\zeta\|=1} R(\zeta) \xrightarrow{p} 0. Q.E.D.$$

The next result is useful for proving Theorem 2.

LEMMA A3: *If* $(R_t, u_t), (t = 1, 2, ...)$ *are i.i.d.,* $E[|u_t|^p|R_t]$ *is bounded for* $p > 0$, $E[u_t|R_t] = 0$, $\text{var}(u_t|R_t) \geq C > 0$, $a_{tT}, (t = 1, ..., T)$ *are random variables depending only* *on* $(R_1, ..., R_T)$, *with*

$$\max_{t\leq T}|a_{tT}| \xrightarrow{p} 0, \ \sum_{t=1}^{T} a_{tT}^2 \operatorname{var}(u_t|R_t) \xrightarrow{p} \Psi > 0,$$

then

$$\sum_{t=1}^{T} a_{tT}u_t \xrightarrow{d} N(0,\Psi).$$

Proof: We proceed by verifying the hypotheses of Lemma 3 of Chamberlain (1986, "Notes on Semiparametric Regression."), denoted L3 henceforth. Note that $\operatorname{var}(u_t|R_t) \leq C$ by $E\left[|u_t|^p|R_t\right]$ bounded, so that $\sum_{t=1}^T a_{tT}^2 \operatorname{var}(u_t|R_t) \leq C\sum_{t=1}^T a_{tT}^2$. Therefore, with probability approaching one, $\sum_{t=1}^T a_{tT}^2 \geq C > 0$. Therefore equation 1) of L3 holds. Equation 2) of L3 is also satisfied, since

$$\left(\max_{t\leq T} a_{tT}^2\right) / \sum_{j=1}^{T} a_{jT}^2 \leq \left(\max_{t\leq T}|a_{tT}|\right)^2 C$$

with probability approaching one. Also, equation (3) of L3 holds, since for $\Delta > 0$,

$$\begin{aligned} E\left[1\left(|u_t| \geq \Delta\right) u_t^2|R_t\right] &= E\left[1\left(\left|\frac{u_t}{\Delta}\right| \geq 1\right)\left|\frac{u_t}{\Delta}\right|^2 \mid R_t\right]\Delta^2 \\ \leq \ & E\left[1\left(\left|\frac{u_t}{\Delta}\right| \geq 1\right)\left|\frac{u_t}{\Delta}\right|^p \mid R_t\right]\Delta^2 \leq E\left[|u_t|^p \mid R_t\right]\Delta^{2-p} \leq C\Delta^{2-p} \end{aligned}$$

which goes to zero as $\Delta \to \infty$. Also, equation 4) of L3 is satisfied by $\operatorname{var}(u_t|R_t)$ bounded away from zero. Let $I_T = 1$ if $\sum_{t=1}^T a_{tT}^2 > 0$, $I_T = 0$ otherwise. Note that for $\sigma_t^2 = \operatorname{var}(u_t|R_t)$

$$\sum_{t=1}^{T} a_{tT}u_t = (1-I_T)\sum_{t=1}^{T} a_{tT}u_t + \left[I_T\sum_{t=1}^{T} a_{tT}u_t / \left(\sum_{t=1}^{T} a_{tT}^2\sigma_t^2\right)^{1/2}\right]\left(\sum_{t=1}^{T} a_{tT}^2\sigma_t^2\right)^{1/2}$$

The first term converges in probability to zero by $I_T = 1$ with probability approaching one. Also $\left(\sum_{t=1}^T a_{tT}^2\sigma_t^2\right)^{1/2} \xrightarrow{p} \Psi^{1/2}$, so by the Slutzky theorem and the conclusion of L3,

$$\sum_{t=1}^{T} a_{tT}u_t \xrightarrow{d} \Psi^{1/2}N(0,1) = N(0,\Psi).Q.E.D.$$

Proof of Theorem 2: We will first prove the asymptotic normality result for LIML. We use an expansion of the first order conditions. For $u(\delta) = y - X\delta$, consider the function

$$\hat{D}(\delta) = -\frac{X'P_Z u(\delta)}{T} + \frac{u(\delta)'P_Z u(\delta)}{u(\delta)'u(\delta)} \frac{X'u(\delta)}{T}.$$

The first-order conditions for $\tilde{\delta}$ are

$$0 = \hat{D}(\tilde{\delta})$$

Then for asymptotic normality it suffices to show that for a nonsingular matrix H, a matrix Σ, and any $\bar{\delta} \xrightarrow{p} \delta_0$,

$$\sqrt{T}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Sigma), \frac{\partial \hat{D}(\bar{\delta})}{\partial \delta} \xrightarrow{p} H.$$

It will then follow in the usual way that

$$\sqrt{T}(\tilde{\delta} - \delta) \xrightarrow{d} N(0, \Lambda), \Lambda = H^{-1}\Sigma H'^{-1}.$$

The first thing we show is convergence in probability of $\partial \hat{D}(\bar{\delta})/\partial \delta$. Let $\bar{u} = u(\bar{\delta}) = y - X\bar{\delta}$. Then differentiating gives

$$\begin{aligned} \frac{\partial \hat{D}}{\partial \delta}(\bar{\delta}) &= \frac{X'P_Z X}{T} - \frac{\bar{u}'P_Z \bar{u}}{\bar{u}'\bar{u}} \frac{X'X}{T} - \frac{X'\bar{u}}{T} \frac{\bar{u}'P_Z X}{\bar{u}'\bar{u}} - \frac{X'P_Z \bar{u}}{\bar{u}'\bar{u}} \frac{\bar{u}'X}{T} + 2\frac{\bar{u}P_Z \bar{u}}{(\bar{u}'\bar{u})^2} \frac{X'\bar{u}\bar{u}'X}{T} \\ &= \frac{X'P_Z X}{T} - \frac{\bar{u}'P_Z \bar{u}}{\bar{u}'\bar{u}} \frac{X'X}{T} + (X'\bar{u}/\bar{u}'\bar{u})\hat{D}(\bar{\delta})' + \hat{D}(\bar{\delta})\bar{u}'X/\bar{u}'\bar{u}. \end{aligned}$$

It then follows in a standard way that $X'\bar{u}/T \xrightarrow{p} \sigma_{Vu}$ and $\bar{u}'\bar{u}/T \xrightarrow{p} \sigma_u^2$. From Lemma A2 we also have $X'P_Z X/T \xrightarrow{p} \alpha\Omega + Q$ and it follows similarly that $\bar{u}'P_Z \bar{u}/T \xrightarrow{p} \alpha\sigma_u^2$. Therefore, $\hat{D}(\bar{\delta}) \xrightarrow{p} 0$, so that

$$\frac{\partial \hat{D}(\bar{\delta})}{\partial \delta} \xrightarrow{p} Q + \alpha\Omega - \alpha(Q + \Omega) = (1 - \alpha)Q = H.$$

Next we consider the behavior of $\sqrt{T}\hat{D}(\delta_0)$. For $\hat{\gamma} = X'u/u'u$ and $\gamma = \sigma_{Vu}/\sigma_u^2$, by $E[\tilde{V}_t u_t] = 0$ and the Lindberg-Levy central limit theorem, $\tilde{V}'u/\sqrt{T}$ is bounded in probability. Also, $\Pi'X'u/\sqrt{T}$ has bounded second moment and so is bounded in probability. Also, by Lemma A2, $u'P_Z u/u'u \xrightarrow{p} \alpha$. Therefore, by the Slutzky theorem,

$$\sqrt{T}(\hat{\gamma} - \gamma)\frac{u'P_Z u}{T} = \frac{X'u - \gamma u'u}{\sqrt{T}} \frac{u'P_Z u}{u'u} = \alpha\frac{(Z\Pi + \tilde{V})'u}{\sqrt{T}} + o_p(1).$$

We then have

$$\begin{aligned}
\sqrt{T}\hat{D}(\delta_0) &= -\frac{X'P_z u}{\sqrt{T}} + \frac{u'P_Z u}{u'u}\frac{X'u}{\sqrt{T}} = -\frac{(X-u\hat{\gamma}')'P_Z u}{\sqrt{T}} \\
&= -\frac{(X-u\gamma')'P_Z u}{\sqrt{T}} + \sqrt{T}(\hat{\gamma}-\gamma)\frac{u'P_Z u}{T} \\
&= -\frac{\Pi' Z' u}{\sqrt{T}} - \frac{\tilde{V}'P_Z u}{\sqrt{T}} + \alpha\frac{(Z\Pi+\tilde{V})'u}{\sqrt{T}} + o_p(1) \\
&= \frac{W'u}{\sqrt{T}} + o_p(1), W = -(1-\alpha)Z\Pi - (P_Z - \alpha I)\tilde{V}.
\end{aligned}$$

Thus, for asymptotic normality of $\sqrt{T}\hat{D}(\delta_0)$ it suffices (by the Slutzky theorem) that $W'u/\sqrt{T} \xrightarrow{d} N(0,\Sigma)$, $\Sigma = (1-\alpha)^2\sigma_u^2 Q + \alpha(1-\alpha)\sigma_u^2\tilde{\Omega}$. Also, by the Cramer-Wold device it suffices to prove that for any vector λ, $\lambda' W' u/\sqrt{T} \xrightarrow{d} N(0, \lambda'\Sigma\lambda)$, or equivalently that the conclusion holds when $G=1$. Without changing notation we will assume that X, $Z\Pi$, and V are vectors, representing $X\lambda$, $Z\Pi\lambda$ and $V\lambda$ respectively. Let

$$a_{tT} = W_t = (1-\alpha)Z_t\Pi + \tilde{V}'Z(Z'Z)^{-1}Z_t - \alpha\tilde{V}_t$$

By hypothesis, $\max_{t\leq T}|Z_t\Pi|/\sqrt{T} \to 0$. Also, by the Markov inequality,

$$\max_{t\leq T}\left|\tilde{V}_t\right|/\sqrt{T} = \left(\max_{t\leq T}\left|\tilde{V}_t\right|^4/T^2\right)^{1/4} \leq \left(\sum_{t=1}^{T}\left|\tilde{V}_t\right|^4/T^2\right)^{1/4} \xrightarrow{p} 0.$$

Also, by the Marcinkiewicz-Zygmund inequality, for $w_{st} = Z_s(Z'Z)^{-1}Z_t'$,

$$E\left[\left|V'Z(Z'Z)^{-1}Z_t\right|^p\right] = E[|\sum_{s=1}^{T}V_s w_{st}|^p] \leq CE[|\sum_{s=1}^{T}V_s^2 w_{st}^2|^{p/2}].$$

As shown above, $Z_t(Z'Z)^{-1}Z_t' \leq 1$, so that

$$\sum_{s=1}^{T} w_{st}^2 = \sum_{s=1}^{T} Z_t(Z'Z)^{-1}Z_s'Z_s(Z'Z)^{-1}Z_t = Z_t(Z'Z)^{-1}Z_t' \leq 1.$$

Hence $\left(\sum_{s=1}^{T} w_{st}^2\right)^{-p/2} \geq 1$, so by Jensen's inequality,

$$\begin{aligned}
E\left[\left|\sum_{s=1}^{T}V_s^2 w_{st}^2\right|^{p/2}\right] &\leq E\left[\left|\sum_{s=1}^{T}V_s^2 w_{st}^2/\sum_{s=1}^{T}w_{st}^2\right|^{p/2}\right] \leq E\left[\sum_{s=1}^{T}\left(|V_s|^2\right)^{p/2} w_{st}^2/\sum_{s=1}^{T}w_{st}^2\right] \\
&\leq \sum_{s=1}^{T}E\left[|V_s|^p\right]w_{st}^2/\sum_{s=1}^{T}w_{st}^2 \leq C.
\end{aligned}$$

Combining the last two equations gives $E\left[|V'Z(Z'Z)^{-1}Z_t|^p\right] \leq C$. Thus, by the Markov inequality, $\sum_{s=1}^{T} |V'Z(Z'Z)^{-1}Z_t|^p /T$ is bounded in probability, and

$$\begin{aligned}
\max_{t\leq T}\left|V'Z(Z'Z)^{-1}Z_t\right|/\sqrt{T} &\leq \left(\sum_{t=1}^{T}\left|V'Z(Z'Z)^{-1}Z_t\right|^p/T^{p/2}\right)^{1/p} \\
&= \left(T^{1-p/2}\sum_{t=1}^{T}\left|V'Z(Z'Z)^{-1}Z_t\right|^p/T\right)^{1/p} \xrightarrow{p} 0.
\end{aligned}$$

Then, by the triangle inequality,

$$\max_{t\leq T}|a_{tT}|/\sqrt{T} \xrightarrow{p} 0.$$

Next, note that by Lemma A2 and the law of large numbers,

$$\begin{aligned}
\tilde{V}'(P_Z-\alpha I)^2\tilde{V}/T &= (1-2a)\tilde{V}'P_Z\tilde{V}/T + \alpha^2\tilde{V}'\tilde{V}/T \xrightarrow{p} \alpha(1-\alpha)\tilde{\Omega}, \\
\Pi'Z'(P_Z-\alpha I)\tilde{V}/T &= (1-\alpha)\Pi'Z'u/T \xrightarrow{p} 0.
\end{aligned}$$

Then by the triangle inequality,

$$\begin{aligned}
W'W/T &= (1-\alpha)^2\Pi'Z'Z\Pi/T + 2(1-\alpha)\Pi'Z'(P_Z-\alpha I)\tilde{V}/T \\
&\quad +\tilde{V}'(P_Z-\alpha I)^2\tilde{V}/T \xrightarrow{p} \Sigma/\sigma_u^2.
\end{aligned}$$

Since $\text{var}(u_t|\tilde{V}) = \sigma_u^2$, we have

$$\sum_{t=1}^{T} a_{tT}^2 \,\text{var}(u_t|\tilde{V}_t)/T = \sigma_u^2 W'W/T \xrightarrow{p} \Sigma.$$

It then follows by the conclusion of Lemma A3 that

$$\frac{W'u}{\sqrt{T}} = \frac{\sum_{t=1}^{T} a_{tT}u_t}{\sqrt{T}} \xrightarrow{d} N(0,\Sigma).$$

The asymptotic normality result for LIML now follows as described above.

For the other estimators, note that $X'u/\sqrt{T} = O_p(\sqrt{T})$ and $X'X/T = O_p(1)$, so that

$$(\hat{\alpha}-\tilde{\alpha})X'u/\sqrt{T} = o_p(1/\sqrt{T})O_p(\sqrt{T}) \xrightarrow{p} 0, (\hat{\alpha}-\tilde{\alpha})X'X/T \xrightarrow{p} 0.$$

Also, note that $\hat{H} = X'P_ZX/T - \hat{\alpha}X'X/T \xrightarrow{p} H$ follows as in the proof of Theorem 1. Therefore, by eq. (6.1),

$$\sqrt{T}(\hat{\delta} - \tilde{\delta}) = -\hat{H}^{-1}(\hat{\alpha} - \tilde{\alpha})X'u/\sqrt{T} + \hat{H}^{-1}(\hat{\alpha} - \tilde{\alpha})X'X/T[\sqrt{T}(\tilde{\delta} - \delta_0)] \xrightarrow{p} 0.$$

Next, to show $T\hat{\Lambda} \xrightarrow{p} \Lambda$, note that by Theorem 1 $\hat{\delta} \xrightarrow{p} \delta_0$, so that $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$ holds by standard arguments, as does $X'\hat{u}/T \xrightarrow{p} \sigma_{Vu}$. Also $\tilde{\alpha} \xrightarrow{p} \alpha$ by Theorem 1, so that $\hat{\alpha} \xrightarrow{p} \alpha$ holds by T and by hypothesis in Theorem 2. Therefore, by Lemma A2,

$$\begin{aligned}
\hat{H}/T \xrightarrow{p} Q + \alpha\Omega - \alpha(Q + \Omega) &= (1 - \alpha)Q, \hat{J}/T \xrightarrow{p} Q + \alpha\Omega - \alpha\sigma_{Vu}\sigma_{Vu}'/\sigma_u^2 = Q + \alpha\tilde{\Omega}, \\
&\hat{\Sigma}/T \xrightarrow{p} \sigma_u^2[(1 - \alpha)(Q + \alpha\tilde{\Omega}) - \alpha(1 - \alpha)Q] \\
&= \sigma_u^2[(1 - \alpha)^2Q + \alpha(1 - \alpha)\tilde{\Omega}].
\end{aligned}$$

The result then follows by the continuous mapping theorem. Q.E.D.

The few results are useful for proving Theorems 3 and 4. Let $\bar{\alpha} = rank(Z)/K$ and $\hat{M} = Z'Z/T$.

LEMMA A4: *If Assumption 3 is satisfied then* $\bar{\alpha} \xrightarrow{p} 1$ *and there is a generalized inverse* $\hat{M}^-$ *with* $\hat{M}^-\hat{M}$ *symmetric and* $\left\|\hat{M}^-\hat{M} - \hat{M}\right\|^2/K \xrightarrow{p} 0$.

Proof: Note that $\bar{\alpha}$ is invariant to nonsingular transformation of Z. Let $Z = Z(E[Z_tZ_t'])^{-1/2}$, so now Assumption 3 is $E[(Z_t'Z_t)^2]/KT \to 0$. Let $\hat{M} = B\Lambda B'$ where B is an orthogonal matrix, Λ is a diagonal matrix of eigenvalues of $\hat{M}$, and let $M^- = B\Lambda^-B'$, where Λ^- is the diagonal g-inverse. Then $\hat{A} = \hat{M}^-M$ is symmetric and idempotent so that

$$\bar{\alpha} = \text{tr}(P_Z)/K = \text{tr}(A)/K = \left\|\hat{M}^-\hat{M}\right\|^2/K.$$

By $E[\left\|\hat{M} - I_K\right\|^2] \le CE[(Z_t'Z_t)^2]/T$ and M,

$$\left\|\hat{M} - I_K\right\|^2/K = O_p\left(E[(Z_t'Z_t)^2]/TK\right) \xrightarrow{p} 0.$$

We also have, for $\hat{A} = \hat{M}^-\hat{M}$

$$\begin{aligned}
\left\|\hat{A} - \hat{M}\right\|^2/K &= \left\|\hat{A} - \hat{A}\hat{M}\right\|^2/K \le \left\|\hat{A}(I - \hat{M})\right\|^2/K \\
&\le \text{tr}((I - \hat{M})\hat{A}^2(I - \hat{M}))/K \le \left\|I_K - \hat{M}\right\|^2/K \xrightarrow{p} 0,
\end{aligned}$$

giving second conclusion. It then follows by T that $\left\|\hat{A}-I_K\right\|^2/K \xrightarrow{p} 0$, so that by $1=\|I_K\|^2/K$,

$$|\bar{\alpha}-1|=\left|\left\|\hat{A}\right\|^2-\|I_K\|^2\right|/K \le \left\|\hat{A}-I_K\right\|^2/K+2\|I_K\|\left\|\hat{A}-I_K\right\|/K \xrightarrow{p} 0.Q.E.D.$$

LEMMA A5: *If Assumption 3 is satisfied then for* $\bar{Q}=\begin{bmatrix}\delta_0'\tilde{Q}\delta_0 & \delta_0'\tilde{Q}\\ \tilde{Q}\delta_0 & \tilde{Q}\end{bmatrix}$,

$$\bar{X}'\bar{X}/T \xrightarrow{p} \bar{\Omega},\ \bar{X}'P_Z\bar{X}/K \xrightarrow{p} \bar{\Omega}+\bar{Q}.$$

Proof: First, assume for the moment that X_t is a scalar, and note that

$$E[(\bar{\Pi}'Z'Z\bar{\Pi}/K-T\bar{\Pi}'E[Z_tZ_t']\bar{\Pi}/K)^2]\le (T/K^2)E[(\bar{\Pi}'Z_t)^4]\longrightarrow 0,$$

so that $\bar{\Pi}'Z'Z\bar{\Pi}/K \xrightarrow{p} \tilde{Q}$ by M and T. Also, $E[\left\|\bar{\Pi}'Z'\bar{V}/K\right\|^2]=(T/K)\left\|\bar{\Pi}'E[Z_tZ_t']\bar{\Pi}/K\right\|(\bar{\Omega}/K)\longrightarrow 0$, so that $\bar{\Pi}'Z'\bar{V}/K \xrightarrow{p} 0$. Applying these results element by element gives the same conclusions for the vector X_t case. By Lemma A1,

$$\bar{V}'P_Z\bar{V}/K-\bar{\alpha}\bar{\Omega}=K^{-1}\{\bar{V}'P_Z\bar{V}-E[\bar{V}'P_Z\bar{V}|\tilde{Z}]\}=O_p(1/\sqrt{K})\xrightarrow{p} 0.$$

Then the second conclusion follows by T, Lemma A4

$$\bar{X}'P_Z\bar{X}=\bar{V}'P_Z\bar{V}+Z\bar{\Pi}'\bar{V}+\bar{V}'Z\bar{\Pi}+\bar{\Pi}'Z'Z\bar{\Pi}.$$

The first conclusion follows similarly. QED.

Proof of Theorem 3: By Lemma A4, when $(T/K)\hat{\alpha}=rank(Z)/K+o_p(1)$ we have

$$\frac{X'P_ZX}{K}-\hat{\alpha}\frac{X'X}{K}=\frac{X'P_ZX}{K}-(T\hat{\alpha}/K)\frac{X'X}{T}\xrightarrow{p}\tilde{Q}+\Omega-1\Omega=\tilde{Q},$$

which is nonsingular by Assumption 1. Also, by Lemma A1,

$$V'P_Zu/K=E[V'P_Zu/K|\tilde{Z}]+O_p(\sqrt{K}/K)=\bar{\alpha}\sigma_{Vu}+o_p(1)\xrightarrow{p}\sigma_{Vu}.$$

and $X'u/T \xrightarrow{p} \sigma_{Vu}$ also holds. The first conclusion then follows from eq. (6.1) and the continuous mapping theorem.

To prove the second conclusion, note that for $a > 0$ and $b \geq 0$, $(a+b)/a \geq 1$,with equality if and only if $b = 0$. By Assumption 3, $\bar{\Omega}$ is nonsingular, so for any vector ζ with $\|\zeta\| = 1$ we have

$$\bar{R}(\zeta) = \frac{\zeta'(\bar{\Omega} + \bar{Q})\zeta}{\zeta'\bar{\Omega}\zeta} \geq 1,$$

with equality when $\zeta'\bar{Q}\zeta = 0$. Let $\hat{R}(\zeta) = (T/K)\zeta'\bar{X}'P_Z\bar{X}\zeta/(\zeta'\bar{X}'\bar{X}\zeta)$. By Lemma A5 $\hat{R}(\zeta) - \bar{R}(\zeta) \xrightarrow{p} 0$, so the rest of the proof follows as in the proof of Theorem 1. Q.E.D.

The next two results are useful for proving Theorem 4.

LEMMA A6: *If Assumptions 3 and 4 are satisfied then*

$$\frac{X'P_Z u}{\sqrt{K}} - \frac{u'P_Z u}{\sqrt{K}}\frac{X'u}{u'u} = \frac{\Pi'Z'u}{\sqrt{K}} + \frac{\tilde{V}'ZZ'u}{T\sqrt{K}} + o_p(1).$$

Proof: We prove the result for scalar X_t, so that the vector X_t result will hold by applying the scalar result to each component. Note that for $\tilde{Z} = [Z_1, ..., Z_n]$, $E\left[u'P_Z u \mid \tilde{Z}\right] \leq CK$ (e.g. see Newey (1997)), so that $u'P_Z u = O_p(K)$. Then by $X'u/u'u \stackrel{def}{=} \hat{\gamma} = \gamma + O_p(1/\sqrt{T})$ for $\gamma = E[V_i u_i]/E[u_i^2]$, we have

$$\frac{u'P_Z u}{\sqrt{K}}(\hat{\gamma} - \gamma) = O_p(K/\sqrt{K})O_p(1/\sqrt{T}) = O_p\left(\sqrt{\frac{K}{T}}\right) \xrightarrow{p} 0$$

by $K/T \to 0$. Then using $X = Z\Pi + V$ we have

$$\frac{X'P_Z u}{\sqrt{K}} - \frac{u'P_Z u}{\sqrt{K}}\hat{\gamma} = \frac{\Pi'Z'u}{\sqrt{K}} + \frac{\tilde{V}'P_Z u}{\sqrt{K}} + o_p(1).$$

Next, note that for $\hat{M} = Z'Z/n$ and $\Delta = \hat{M}^{-} - I_K$,

$$\begin{aligned}
& E\left[\left(\frac{\tilde{V}'P_Z u}{\sqrt{K}} - \frac{\tilde{V}'ZZ'u}{T\sqrt{K}}\right)^2 \mid \tilde{Z}\right] = E\left[\left(\tilde{V}'Z\Delta Z'u/T\sqrt{K}\right)^2 \mid \tilde{Z}\right] \\
= & \sum_i \sum_j \sum_k \sum_\ell E\left[\tilde{V}_i \tilde{V}_j u_k u_\ell \mid \tilde{Z}\right] Z_i'\Delta Z_j Z_k'\Delta Z_\ell / KT^2 \\
= & \sum_i E\left[\tilde{V}_i^2 u_i^2 \mid Z_i\right](Z_i'\Delta Z_i)^2/KT^2 + \sum_{i \neq j} E\left[\tilde{V}_i^2 \mid Z_i\right] E\left[u_j^2 \mid Z_j\right](Z_i'\Delta Z_j)^2/KT^2 \\
\leq & C\sum_{i,j} Z_i'\Delta(Z_j Z_j'/n)\Delta Z_i/TK \leq C\sum_{i=1}^n Z_i'\Delta\hat{M}\Delta Z_i/TK = C\,\mathrm{tr}\left(\left(\Delta\hat{M}\right)^2\right)/K.
\end{aligned}$$

Since P_Z is invariant to is invariant to the choice of generalized inverse, the above inequalities hold for any generalized inverse. By Lemma A4 there is a generalized inverse with ΔM symmetric, so that Let B be an orthogonal matrix and Λ a diagonal matrix of eigenvalues with $\hat{M} = B'\Lambda B$. Let Λ^- denote the diagonal matrix of inverse eigenvalues where they are non-zero and zero otherwise. Then $\hat{M}^-\hat{M} = B'\Lambda^-\Lambda B \leq I$ is symmetric and p.s.d. so that,

$$\begin{aligned} \operatorname{tr}\left(\left(\Delta\hat{M}\right)^2\right)/K &= \left\|\Delta\hat{M}\right\|^2/K = \left\|(I-\hat{M})\hat{M}^-\hat{M}\right\|^2/K \\ &\leq \left\|I-\hat{M}\right\|^2/K = O_p\left(E\left[(Z_t'Z_t)^2\right]/KT\right) \xrightarrow{p} 0, \end{aligned}$$

where the last equality follows as in Newey (1997). The conclusion then follows by T and CM. Q.E.D..

For the next result let W_{in}, U_{in} denote $m \times 1$ random vectors, where m can depend on n. Also, let a_n denote an $m \times 1$ vector of constants. The following Lemma is proved in Newey (2004).

LEMMA A6: *If* $\Psi = E[W_{in}W_{in}']$ *exists,* $E[U_{in}U_{in}'] = I_m$, $E[W_{in}] = E[U_{in}] = 0$, $na_n'a_n \to H$, $n^2\operatorname{tr}(\Psi) \to \Lambda^*$, $n^3a_n'\Psi a_n \to 0$, $\operatorname{tr}(\Psi^2)/[\operatorname{tr}(\Psi)]^2 \to 0$, $nE\left[|a_n'U_{in}|^4\right] \to 0$, $n^{-1}E[|W_{1n}'U_{2n}|^4]/\operatorname{tr}(\Psi)^2 \to 0$, *then*

$$\sum_{i=1}^{n} a_n'U_{in} + \sum_{i,j=1}^{n} W_{in}'U_{jn} \xrightarrow{d} N(0, H+\Lambda^*)$$

Proof of Theorem 4: We proceed similarly to the proof of Theorem 2, first proving the result for the LIML estimator, replacing $\hat{D}(\delta)$ there with

$$\tilde{D}(\delta) = -\frac{X'P_Zu(\delta)}{K} + \frac{u(\delta)'P_Zu(\delta)}{u(\delta)'u(\delta)}\frac{X'u(\delta)}{K}.$$

We will first show asymptotic normality of $\sqrt{K}\tilde{D}(\delta_0)$. We do this for the scalar case; the Cramer-Wold device implies the general case. Apply Lemma A6 with $i = t$, $n = T$, $m = K$, $a_n = \Pi\sigma_u/\sqrt{K}$, $W_{in} = \sigma_u Z_t\tilde{V}_t/T\sqrt{K}$, and $U_{in} = Z_tu_t/\sigma_u$. Then $\Psi = I_K\sigma_u^2\tilde{\Omega}/T^2K$, $tr(\Psi) = \sigma_u^2\tilde{\Omega}/T^2$,

$$na_n'a_n = \sigma_u^2T\Pi'\Pi/K \longrightarrow \sigma_u^2\tilde{Q},\ n^2tr(\Psi) = \sigma_u^2\tilde{\Omega}.$$

Also,

$$\operatorname{tr}(\Psi^2)/\operatorname{tr}(\Psi)^2 = \{\sigma_u^4\tilde{\Omega}^2 tr(I_K^2)/T^4K^2\}/\{T^4\sigma_u^4\tilde{\Omega}^2\} = 1/K \longrightarrow 0,$$

and by $|\Pi' Z_i| \leq C$ and boundedness of fourth conditional moments, for $A = E[\|Z_1\|^2 Z_1 Z_1']$,

$$\begin{aligned} nE\left[|a_n' U_{in}|^4\right] &= TE\left[|\Pi' Z_i|^4 u_i^4\right]/K^2 \longrightarrow 0, \\ n^{-1}E\left[|W_{1n}' U_{2n}|^4\right]/(\operatorname{tr}(\Psi))^2 &\leq CT^{-1}\{E\left[(Z_1'Z_2)^4 \tilde{V}_1^4 u_2^4\right]/T^4K^2\}/T^{-4} \\ &\leq CE\left[\|Z_1\|^2 Z_1'(\|Z_2\|^2 Z_2 Z_2')Z_1\right]/TK^2 \\ &= CE\left[\|Z_1\|^2 Z_1' A Z_1\right]/TK^2 = CE\left[tr(A\|Z_1\|^2 Z_1 Z_1')\right]/TK^2 \\ &= Ctr(A^2)/TK^2 \longrightarrow 0. \end{aligned}$$

Then by the Lemmas A5 and A6 and the Slutzky Theorem,

$$\begin{aligned} \sqrt{K}\tilde{D}(\delta_0) &= \frac{X'P_Z u}{\sqrt{K}} - \frac{u'P_Z u}{\sqrt{K}}\frac{X'u}{u'u} = \frac{\Pi' Z' u}{\sqrt{K}} + \frac{\tilde{V}' Z Z' u}{T\sqrt{K}} + o_p(1) \\ &= \sum_{i=1}^{n} a_n' U_{in} + \sum_{i,j=1}^{n} W_{in}' U_{jn} + o_p(1) \xrightarrow{d} N(0, \sigma_u^2\tilde{Q} + \sigma_u^2\tilde{\Omega}). \end{aligned}$$

It also follows similarly to the proof of Theorem 2 that for any $\bar{\delta} \xrightarrow{p} \delta_0$,

$$\partial\tilde{D}(\bar{\delta})/\partial\delta \xrightarrow{p} \tilde{Q}.$$

The remainder of the proof follows similarly to the proof of Theorem 2.

For the other estimators, note that $X'u/\sqrt{K} = O_p(T/\sqrt{K})$ and $X'X/K = O_p(T/K)$, so that

$$(\hat{\alpha}-\tilde{\alpha})X'u/\sqrt{K} = o_p(\sqrt{K}/T)O_p(T/\sqrt{K}) \xrightarrow{p} 0, (\hat{\alpha}-\tilde{\alpha})X'X/K = o_p(\sqrt{K}/T)O_p(T/K) \xrightarrow{p} 0.$$

Also, note that $\hat{H} = X'P_Z X/K - \hat{\alpha}X'X/K \xrightarrow{p} \tilde{Q}$ follows as in the proof of Theorem 1. Therefore, by eq. (6.1),

$$\sqrt{K}(\hat{\delta} - \tilde{\delta}) = -\hat{H}^{-1}(\hat{\alpha} - \tilde{\alpha})X'u/\sqrt{K} + \hat{H}^{-1}(\hat{\alpha} - \tilde{\alpha})(X'X/K)[\sqrt{K}(\tilde{\delta} - \delta_0)] \xrightarrow{p} 0.$$

Next, to show $K\hat{\Lambda} \xrightarrow{p} \tilde{\Lambda}$, note that by Theorem 3 $\hat{\delta} \xrightarrow{p} \delta_0$, so that $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$ holds by standard arguments, as does $X'\hat{u}/T \xrightarrow{p} \sigma_{Vu}$. Also $T\tilde{\alpha}/K \xrightarrow{p} 1$ by Theorem 3 and by hypothesis of Theorem 3, $T(\hat{\alpha} - \tilde{\alpha})/K = (T/K)o_p(\sqrt{K}/T) = o_p(1/\sqrt{K}) \xrightarrow{p} 0$, so that $T\hat{\alpha}/K \xrightarrow{p} 1$ holds. Therefore, it follows similarly to the proof of Theorem 3 that

$$\hat{H}/K \xrightarrow{p} \tilde{Q}, \hat{J}/K \xrightarrow{p} \tilde{Q} + \Omega - \sigma_{Vu}\sigma_{Vu}'/\sigma_u^2 = \tilde{Q} + \tilde{\Omega}.$$

Also, $\check{\alpha} \xrightarrow{p} 0$, so that $\hat{\Sigma}/K \xrightarrow{p} \sigma_u^2[(1-0)(\tilde{Q}+\tilde{\Omega}) + 0\tilde{Q}] = \sigma_u^2(\tilde{Q}+\tilde{\Omega})$. The conclusion follows by the continuous mapping theorem. Q.E.D.

Proof of Theorem 5: As in the proof of Theorem 4 we have

$$\tilde{X}(\delta_0)'P_Z u/\sqrt{K} \xrightarrow{d} N(0, \sigma_u^2(\tilde{Q}+\tilde{\Omega})).$$

Also, $X'u/u'u \xrightarrow{p} \gamma = \sigma_{Vu}/\sigma_u^2$ and by Lemma A1, $X'P_Z u/K - \bar{\alpha}\sigma_{Vu} \xrightarrow{p} 0$, $u'P_Z u/m \xrightarrow{p} \sigma_u^2$, $X'P_Z X/m \xrightarrow{p} \tilde{Q} + E[V_t V_t']$. Therefore

$$\tilde{X}(\delta_0)'P_Z \tilde{X}(\delta_0)/m \xrightarrow{p} \tilde{Q} + E[V_t V_t'] - \sigma_{Vu}\gamma' - \gamma\sigma_{Vu}' + \sigma_u^2\gamma\gamma' = \tilde{Q} + \tilde{\Omega}.$$

The conclusion then follows by the Slutzky Theorem in the usual way. Q.E.D.

References

Anderson, T.W. and T. Sawa (1979): "Evaluation of the Distribution Function of the Two Stage Least Squares Estimator", *Econometrica,* 47, 163-182.

Angrist, J. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records;" *American Economic Review,* 80, 313-335.

Angrist, J. and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics, 106*, 979–1014.

Angrist, J.D., G.W. Imbens and A. Krueger (1999): "Jackknife Instrumental Variables Estimation", *Journal of Applied Econometrics,14*, 57-67.

Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica,* 63, 657-681.

Bekker, P. and F. Kleibegen (2003): "Finite Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic," *Econometric Theory 19*, 744-753.

Bound, J., D. Jaeger, and R. Baker (1996): "Problems with Instrumental Variables Estimation when the Correlation Between Instruments and the Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association, 90,* 443-450.

Cruz, L.M. and M.J. Moreira (2003): "On the Validity of Econometric Techniques With Weak Instruments: Inference on Returns to Education Using Compulsory Schooling Laws," mimeo.

Davidson, R. and J.G. MacKinnon (2004): "The Case Against Jive," mimeo.

Donald, S. G. and W. K. Newey (2001) "Choosing the Number of Instruments" *Econometrica.*

Fuller, W.A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica* 45, 939-954.

Hahn, J. and J. Hausman (2002) "A New Specification Test for the Validity of Instrumental Variables", *Econometrica 70,* 163-189.

Hahn, J., J.A. Hausman, and G.M. Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal,*

Hahn, J. and A. Inoue (2002): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," *Econometric Reviews* 21, 309-336.

Kleibegen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica* 70, 1781-1803.

Moreira, M. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica* 71, 1027-1048.

Morimune, K. (1983): "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability is Large Compared with the Sample Size," *Econometrica,* 51, 821-841.

Nagar, A.L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", *Econometrica*, 27, 575–595.

Rothenberg, T.J. (1983): "Asymptotic Properties of Some Estimators in Structural Models", in *Studies in Econometrics, Time Series and Multivariate Statistics*, edited by Karlin, S. T Amemiya and L.A. Goodman. New York : Academic Press.

Rothenberg, T.J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in Griliches, Z and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. 4,* New York: Elsevier.

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments", *Econometrica, 65*, 557-586.

Table 2: Asymptotic Distribution of LIML and FULLER

			LIML			FULLER		
rho	K	mu^2	Median	IQR	p; .05	Median	IQR	p; .05
0.5	1	1	0.202	1.217	0.024	0.381	0.466	0.182
0.5	1	2	0.090	0.955	0.031	0.268	0.493	0.132
0.5	1	4	0.025	0.704	0.039	0.149	0.463	0.086
0.5	1	8	0.002	0.494	0.043	0.067	0.396	0.061
0.5	1	16	0.000	0.344	0.043	0.031	0.311	0.053
0.5	1	32	0.000	0.241	0.042	0.016	0.229	0.048
0.5	1	64	0.000	0.169	0.045	0.008	0.165	0.048
0.5	1	128	0.000	0.119	0.048	0.004	0.118	0.049
0.5	2	1	0.238	1.305	0.036	0.361	0.599	0.094
0.5	2	2	0.124	1.032	0.040	0.260	0.571	0.084
0.5	2	4	0.039	0.753	0.043	0.149	0.504	0.071
0.5	2	8	0.005	0.518	0.045	0.068	0.415	0.061
0.5	2	16	0.001	0.353	0.044	0.032	0.319	0.054
0.5	2	32	0.000	0.244	0.042	0.016	0.232	0.048
0.5	2	64	0.000	0.171	0.045	0.008	0.167	0.048
0.5	2	128	0.000	0.120	0.047	0.004	0.119	0.049
0.5	4	1	0.284	1.395	0.055	0.362	0.746	0.092
0.5	4	2	0.168	1.135	0.054	0.267	0.683	0.085
0.5	4	4	0.064	0.828	0.053	0.157	0.572	0.075
0.5	4	8	0.012	0.563	0.050	0.072	0.452	0.064
0.5	4	16	0.001	0.370	0.045	0.032	0.334	0.055
0.5	4	32	0.000	0.250	0.043	0.016	0.238	0.049
0.5	4	64	0.000	0.172	0.045	0.008	0.168	0.048
0.5	4	128	0.000	0.121	0.047	0.004	0.119	0.049
0.5	8	1	0.327	1.475	0.077	0.376	0.890	0.108
0.5	8	2	0.216	1.253	0.073	0.286	0.813	0.099
0.5	8	4	0.100	0.935	0.066	0.175	0.674	0.086
0.5	8	8	0.024	0.633	0.056	0.080	0.511	0.070
0.5	8	16	0.002	0.402	0.048	0.033	0.362	0.056
0.5	8	32	0.000	0.261	0.043	0.016	0.249	0.049
0.5	8	64	0.000	0.177	0.045	0.008	0.172	0.048
0.5	8	128	0.000	0.122	0.047	0.004	0.120	0.049
0.5	16	1	0.369	1.552	0.098	0.398	1.036	0.121
0.5	16	2	0.269	1.370	0.091	0.316	0.957	0.111
0.5	16	4	0.148	1.074	0.080	0.205	0.804	0.096
0.5	16	8	0.045	0.739	0.066	0.096	0.604	0.078
0.5	16	16	0.005	0.457	0.052	0.035	0.410	0.060
0.5	16	32	0.000	0.282	0.045	0.015	0.268	0.050
0.5	16	64	0.000	0.185	0.045	0.008	0.180	0.048
0.5	16	128	0.000	0.125	0.048	0.004	0.123	0.049
0.5	32	1	0.399	1.601	0.114	0.416	1.163	0.131
0.5	32	2	0.320	1.473	0.107	0.349	1.100	0.122
0.5	32	4	0.204	1.217	0.095	0.244	0.952	0.108
0.5	32	8	0.083	0.872	0.077	0.125	0.726	0.087
0.5	32	16	0.015	0.545	0.059	0.044	0.489	0.066
0.5	32	32	0.000	0.320	0.047	0.016	0.304	0.052
0.5	32	64	0.000	0.199	0.046	0.007	0.194	0.048
0.5	32	128	0.000	0.130	0.048	0.004	0.129	0.049

Notes: 500,000 reps., 245.07 secs.

Table 3: Asymptotic Distribution of LIML and FULLER

rho	K	mu^2	LIML Median	LIML IQR	LIML p; .05	FULLER Median	FULLER IQR	FULLER p; .05
0.5	1	1	0.201	1.221	0.024	0.380	0.467	0.182
0.5	1	2	0.092	0.956	0.031	0.268	0.494	0.132
0.5	1	3	0.044	0.804	0.035	0.195	0.485	0.102
0.5	1	4	0.024	0.703	0.038	0.149	0.463	0.085
0.5	1	5	0.011	0.630	0.040	0.116	0.444	0.074
0.5	1	6	0.005	0.573	0.042	0.094	0.426	0.069
0.5	1	7	0.004	0.529	0.042	0.079	0.410	0.064
0.5	1	8	0.001	0.494	0.043	0.066	0.396	0.061
0.5	1	9	0.000	0.463	0.043	0.058	0.381	0.059
0.5	1	10	0.001	0.441	0.044	0.052	0.371	0.059
0.5	1	11	-0.001	0.418	0.043	0.045	0.359	0.057
0.5	1	12	0.000	0.399	0.043	0.042	0.347	0.056
0.5	1	13	0.000	0.383	0.043	0.038	0.338	0.055
0.5	1	14	0.000	0.368	0.043	0.036	0.328	0.055
0.5	1	15	0.001	0.355	0.043	0.034	0.319	0.054
0.5	1	16	0.000	0.344	0.042	0.031	0.311	0.052
0.5	1	17	0.000	0.332	0.042	0.030	0.302	0.053
0.5	1	18	0.000	0.323	0.042	0.028	0.296	0.052
0.5	1	19	0.000	0.314	0.042	0.026	0.289	0.051
0.5	1	20	0.000	0.306	0.042	0.024	0.283	0.051
0.5	1	21	0.000	0.298	0.042	0.024	0.276	0.050
0.5	1	22	0.000	0.292	0.043	0.023	0.272	0.051
0.5	1	23	0.000	0.286	0.042	0.021	0.267	0.050
0.5	1	24	0.000	0.279	0.042	0.021	0.261	0.050
0.5	1	25	0.000	0.274	0.042	0.020	0.257	0.050
0.5	1	26	0.000	0.268	0.042	0.019	0.252	0.049
0.5	1	27	0.000	0.263	0.042	0.019	0.248	0.049
0.5	1	28	0.000	0.257	0.041	0.018	0.243	0.048
0.5	1	29	0.000	0.253	0.042	0.017	0.240	0.049
0.5	1	30	0.001	0.249	0.042	0.017	0.236	0.049
0.5	1	31	-0.001	0.245	0.042	0.015	0.233	0.048
0.5	1	32	0.000	0.241	0.042	0.015	0.230	0.048
0.5	1	36	0.000	0.228	0.043	0.014	0.218	0.048
0.5	1	40	-0.001	0.215	0.043	0.012	0.207	0.048
0.5	1	44	0.000	0.205	0.043	0.011	0.197	0.048
0.5	1	48	0.000	0.196	0.043	0.010	0.190	0.048
0.5	1	52	0.000	0.188	0.044	0.009	0.182	0.048
0.5	1	56	0.000	0.181	0.044	0.009	0.176	0.048
0.5	1	60	0.000	0.175	0.045	0.008	0.171	0.048
0.5	1	64	0.000	0.169	0.045	0.008	0.165	0.048
0.5	1	68	0.000	0.164	0.046	0.007	0.161	0.048
0.5	1	72	0.000	0.160	0.046	0.007	0.156	0.048
0.5	1	76	0.000	0.156	0.046	0.006	0.153	0.048
0.5	1	80	0.000	0.152	0.046	0.006	0.149	0.048
0.5	1	84	0.000	0.147	0.046	0.006	0.145	0.048
0.5	1	88	0.000	0.144	0.046	0.006	0.142	0.048
0.5	1	92	0.000	0.141	0.047	0.005	0.139	0.049
0.5	1	96	0.000	0.138	0.046	0.005	0.136	0.048
0.5	1	100	0.000	0.135	0.046	0.005	0.133	0.048
0.5	1	104	0.000	0.133	0.047	0.005	0.131	0.049
0.5	1	108	0.000	0.130	0.047	0.005	0.129	0.049
0.5	1	112	0.000	0.128	0.046	0.004	0.126	0.048
0.5	1	116	0.000	0.126	0.047	0.004	0.124	0.048
0.5	1	120	0.000	0.123	0.047	0.004	0.122	0.048
0.5	1	124	0.000	0.121	0.047	0.004	0.120	0.049
0.5	1	128	0.000	0.119	0.047	0.004	0.118	0.048

Notes: 500,000 reps., 159.49 secs.

Table Four

Simulation Results, $K = 1, P'z'zP/s2 = 10$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	-0.0042	0.4375	0.0367	-0.0014	0.4403	0.0524
LIML	-0.0042	0.4375	0.0367	-0.0014	0.4403	0.0524
LIML-Bekker			0.0367			0.0524
Fuller (1)	0.0405	0.3745	0.0410	0.0621	0.3611	0.0623
Fuller (1) - Bekker			0.0348			0.0555
Kleibergen			0.0543			0.0500
Kleibergen-Bound			0.0531			0.0490

Simulation Results, $K = 3, P'z'zP/s2 = 10$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0725	0.3848	0.0591	0.1048	0.3707	0.0967
LIML	0.0048	0.4811	0.0491	-0.0047	0.4877	0.0635
LIML-Bekker			0.0413			0.0534
Fuller (1)	0.0453	0.4128	0.0527	0.0590	0.3938	0.0764
Fuller (1)-Bekker			0.0396			0.0578
Kleibergen			0.0524			0.0533
Kleibergen-Bound			0.0496			0.0488

Simulation Results, $K = 5, P'z'zP/s2 = 10$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.1277	0.3454	0.0882	0.1862	0.3254	0.1680
LIML	0.0109	0.5233	0.0636	0.0104	0.5042	0.0844
LIML-Bekker			0.0433			0.0641
Fuller (1)	0.0521	0.4449	0.0692	0.0716	0.4126	0.0997
Fuller (1) - Bekker			0.0408			0.0678
Kleibergen			0.0522			0.0524
Kleibergen-Bound			0.0465			0.0464

Simulation Results, $K = 10, P'z'zP/s2 = 10$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.1938	0.2927	0.1670	0.2914	0.2674	0.3512
LIML	0.0110	0.5987	0.0886	0.0171	0.5590	0.1127
LIML-Bekker			0.0438			0.0663
Fuller (1)	0.0499	0.5075	0.0959	0.0775	0.4492	0.0.1289
Fuller (1)-Bekker			0.0417			0.0698
Kleibergen			0.0549			0.0564
Kleibergen-Bound			0.0439			0.0465

Table Five

Simulation Results, $K = 1, \Pi' z' z \Pi / \sigma^2 = 20$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	-0.0006	0.3048	0.0391	0.0038	0.3063	0.0505
LIML	-0.0006	0.3048	0.0391	0.0038	0.3063	0.0505
LIML-Bekker			0.0391			0.0505
Fuller (1)	0.0198	0.2836	0.0552	0.1100	0.2248	0.0982
Fuller (1) - Bekker			0.0401			0.0770
Kleibergen			0.0527			0.0499
Kleibergen-Bound			0.0509			0.0488

Simulation Results, $K = 3, \Pi' z' z \Pi / \sigma^2 = 20$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0391	0.2822	0.0536	0.0627	0.2831	0.0874
LIML	0.0012	0.3156	0.0461	0.0059	0.3218	0.0624
LIML-Bekker			0.0404			0.0561
Fuller (1)	0.0207	0.2941	0.0489	0.0358	0.2953	0.0734
Fuller (1)-Bekker			0.0391			0.0618
Kleibergen			0.0488			0.0533
Kleibergen-Bound			0.0454			0.0487

Simulation Results, $K = 5, \Pi' z' z \Pi / \sigma^2 = 20$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0714	0.2646	0.0733	0.1033	0.2585	0.1161
LIML	0.0019	0.3361	0.0555	-0.0039	0.3314	0.0630
LIML-Bekker			0.0419			0.0540
Fuller (1)	0.0212	0.3090	0.0580	0.0267	0.3028	0.0726
Fuller (1) - Bekker			0.0410			0.0585
Kleibergen			0.0517			0.0536
Kleibergen-Bound			0.0458			0.0472

Simulation Results, $K = 10, \Pi' z' z \Pi / \sigma^2 = 20$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.1288	0.2406	0.1267	0.1886	0.2268	0.2524
LIML	0.0020	0.3745	0.0714	-0.0013	0.3706	0.0807
LIML-Bekker			0.0432			0.0573
Fuller (1)	0.0208	0.3446	0.0748	0.0287	0.3357	0.0921
Fuller (1)-Bekker			0.0427			0.0620
Kleibergen			0.0554			0.0563
Kleibergen-Bound			0.0424			0.0457

Table Six

Simulation Results, $K = 1, \Pi' z' z \Pi / \sigma^2 = 35$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0028	0.2317	0.0470	-0.0017	0.2266	0.0449
LIML	0.0028	0.2317	0.0470	-0.0017	0.2266	0.0449
LIML-Bekker			0.0470			0.0449
Fuller (1)	0.0147	0.2223	0.0492	0.0153	0.2157	0.0501
Fuller (1) - Bekker			0.0451			0.0491
Kleibergen			0.0564			0.0517
Kleibergen-Bound			0.0554			0.0499

Simulation Results, $K = 3, \Pi' z' z \Pi / \sigma^2 = 35$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0236	0.2209	0.0516	0.0352	0.2206	0.0667
LIML	0.0009	0.2362	0.0479	0.0023	0.2401	0.0515
LIML-Bekker			0.0433			0.0486
Fuller (1)	0.0123	0.2277	0.0487	0.0187	0.2274	0.0573
Fuller (1)-Bekker			0.0406			0.0517
Kleibergen			0.0500			0.0533
Kleibergen-Bound			0.0467			0.0497

Simulation Results, $K = 5, \Pi' z' z \Pi / \sigma^2 = 35$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0406	0.2125	0.0647	0.0620	0.2101	0.0914
LIML	-0.0029	0.2426	0.0516	-0.0012	0.2406	0.0522
LIML-Bekker			0.0434			0.0473
Fuller (1)	0.0088	0.2340	0.0534	0.0160	0.2279	0.0588
Fuller (1) - Bekker			0.0421			0.0505
Kleibergen			0.0558			0.0533
Kleibergen-Bound			0.0493			0.0471

Simulation Results, $K = 10, \Pi' z' z \Pi / \sigma^2 = 35$

	$\rho = .4$			$\rho = .6$		
	Med. Bias	**IQR**	**Size**	**Med. Bias**	**IQR**	**Size**
2SLS	0.0846	0.1940	0.1034	0.1263	0.1866	0.1784
LIML	0.0003	0.2531	0.0628	0.0024	0.2568	0.0643
LIML-Bekker			0.0427			0.0481
Fuller (1)	0.0122	0.2430	0.0646	0.0198	0.2425	0.0698
Fuller (1)-Bekker			0.0415			0.0512
Kleibergen			0.0547			0.0536
Kleibergen-Bound			0.0420			0.0413

FIGURE 3

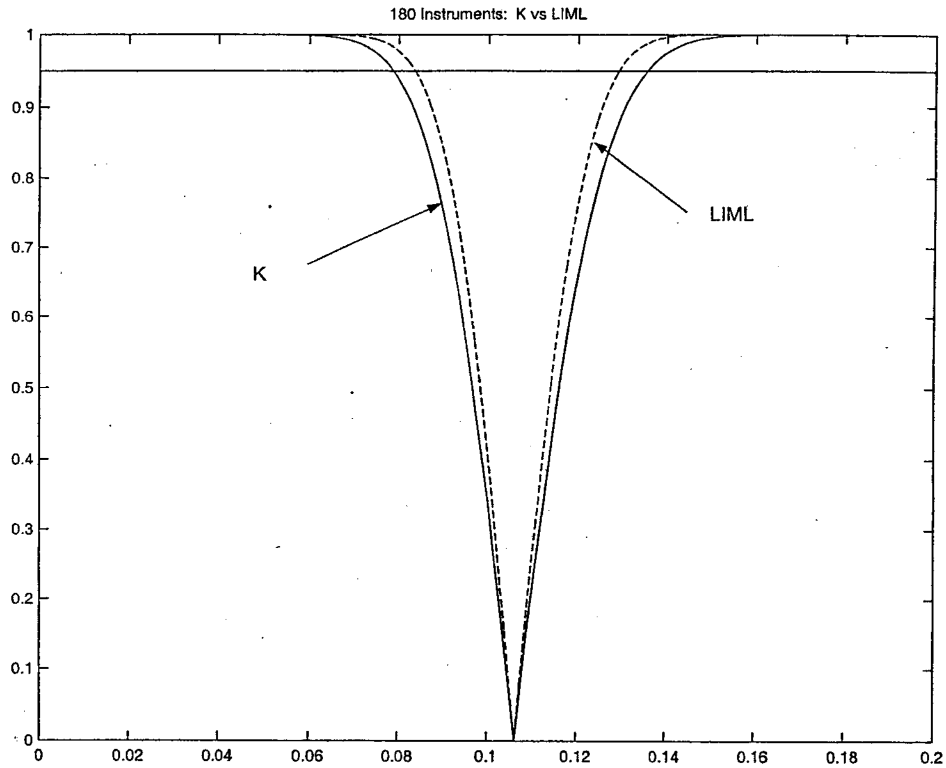


FIGURE 4

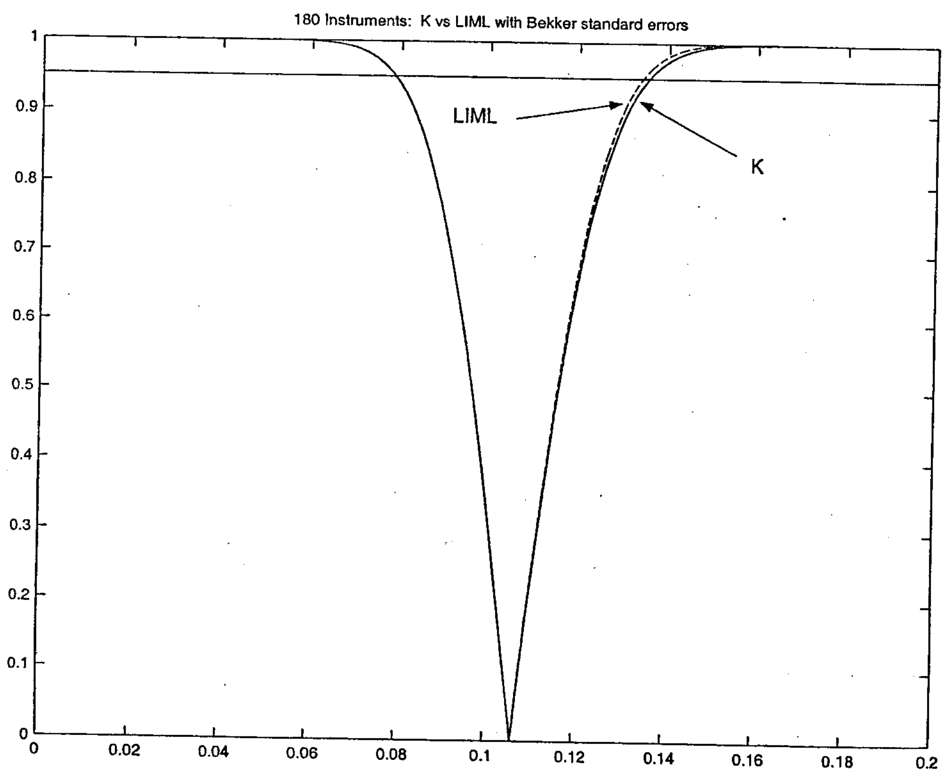


FIGURE 1

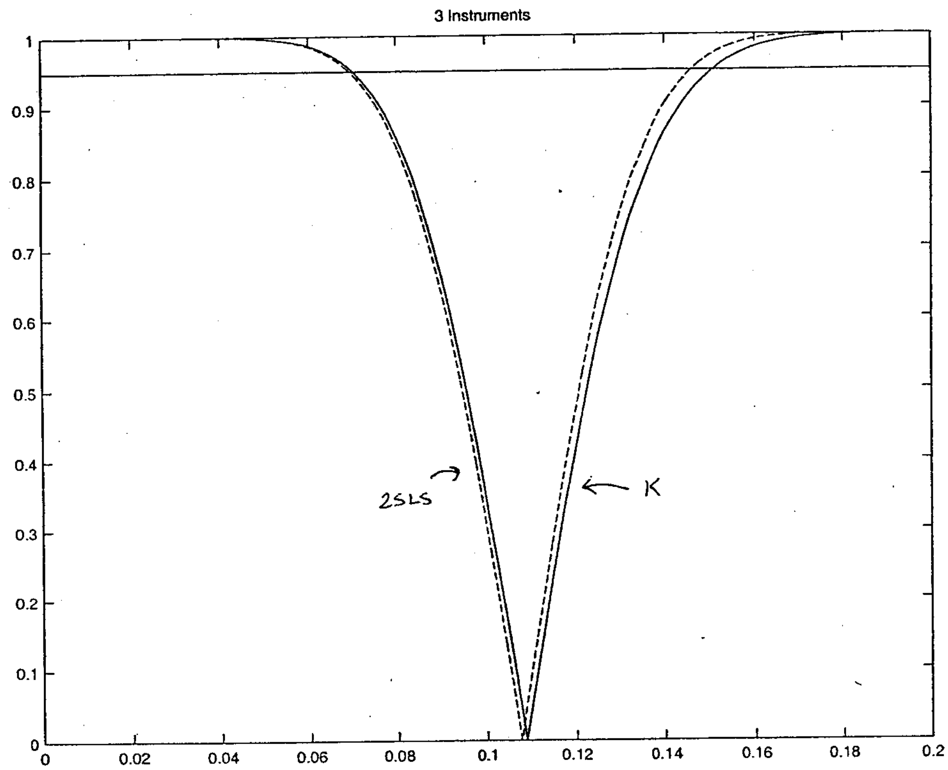


FIGURE 2

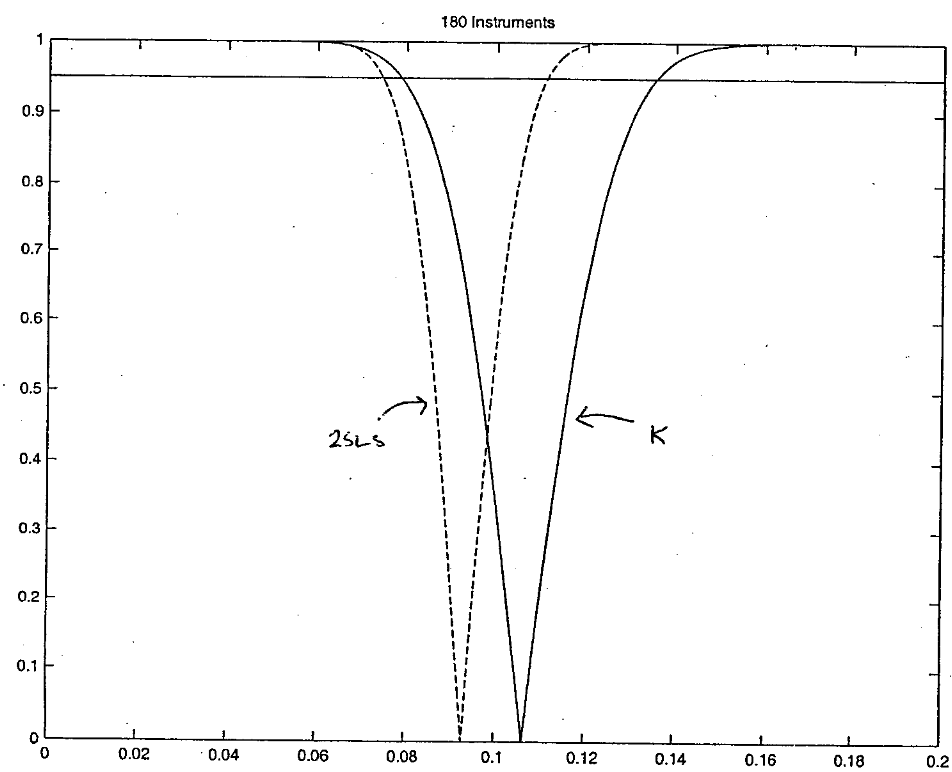


FIGURE 5

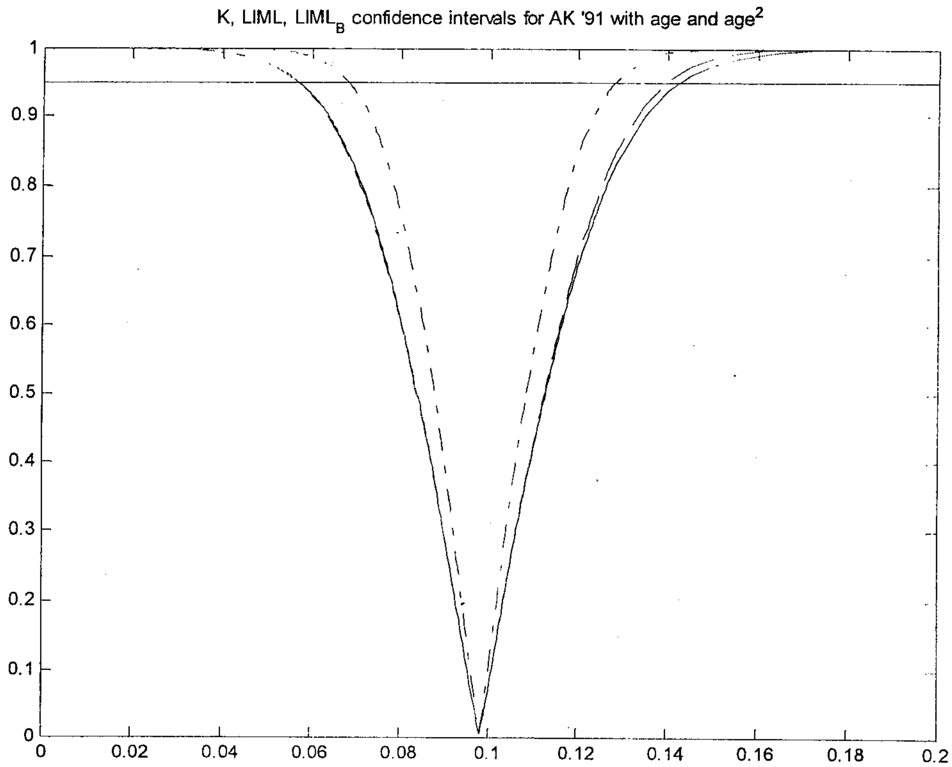
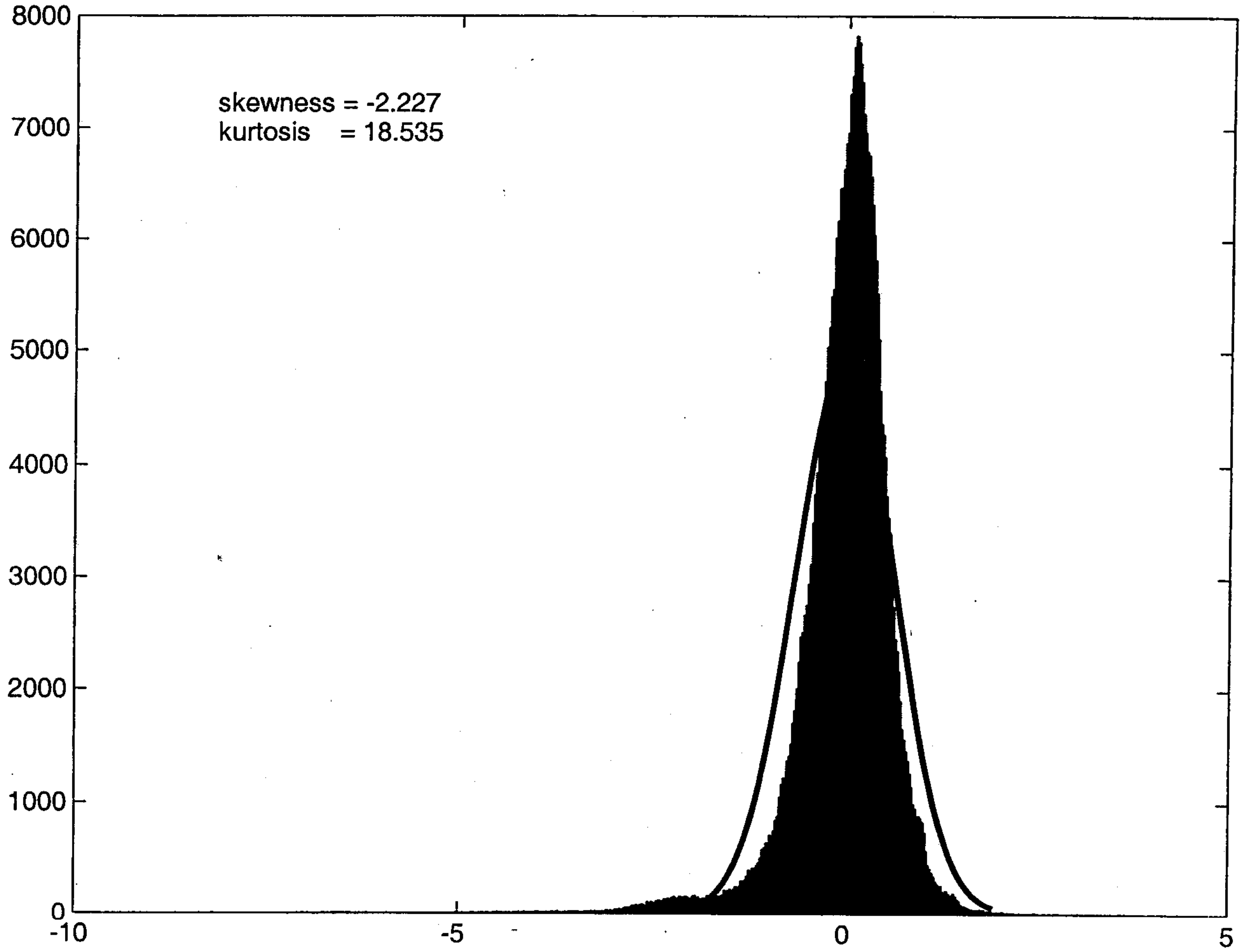

K, LIML, LIML$_B$ confidence intervals for AK '91 with age and age^2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2

FIGURE 6


8000
7000
6000
5000
4000
3000
2000
1000
0
-10
-5
0
5
skewness = -2.227
kurtosis = 18.535